

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2002, 09:56:54 ; Search time 38.59 seconds
(without alignments)
3373.374 Million cell updates/sec

Title: US-09-822-682-2
Perfect score: 6605
Sequence: 1 MWRLVLLALWSPSTQAGH.....FVFSQEMVYFSLKVECRDI 1172

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6605	100.0	1172	21 AAB19677	Human thrombospondin
2	6570	99.5	1172	21 AAB00043	Human thrombospondin
3	4228	64.0	1170	22 AAB90800	Human shear stress
4	4228	64.0	1170	22 AAB74450	Human variant thro
5	4226	64.0	1152	21 AAB00042	Human thrombospondin
6	3251	49.2	571	22 AAG75130	Human colon cancer
7	3236	49.0	571	21 AAB43386	Human cancer assoc
8	2094	31.7	731	22 AAU02913	Angiotensin conver
9	1990	30.1	757	21 AAB00044	Human cartilage ol
10	1946.5	29.5	762	22 AAM93335	Human polypeptide,
11	1899	28.8	899	15 AAB56248	Xenopus thrombospo

12	1890	28.6	961	15 AAB56249	Human thrombospondin
13	1805	27.3	1024	22 AAB56249	Drosophila melanog
14	1428	21.6	444	21 AAB58933	Breast and ovarian
15	1428	21.6	444	22 AAG73734	Human colon cancer
16	1359	20.6	555	22 AAU02914	Angiotensin conver
17	1302	19.7	776	22 AAM79078	Human protein SEQ
18	1193	18.1	541	22 AAB74451	Human variant thro
19	1123	17.0	546	22 AAU02915	Angiotensin conver
20	1012.5	15.3	300	21 AAB00041	Human COMP/TSP-2 c
21	880	13.3	239	14 AAB0823	Human thrombospondin
22	850	12.9	459	22 AAU02916	Angiotensin conver
23	817	12.4	441	19 AAM40288	Human concatameris
24	783	11.9	441	20 AAY06183	Thrombospondin I f
25	779	11.8	466	21 AAB43602	Human cancer assoc
26	750.5	11.4	622	22 AAM80062	Human protein SEQ
27	748	11.3	218	19 AAM40287	Human TSP1 protein
28	713	10.8	218	20 AAY06182	Thrombospondin I f
29	645	9.8	157	21 AAB08133	Amino acid sequenc
30	612	9.3	151	21 AAB53711	Human colon cancer
31	513.5	7.8	242	21 AAB00040	Human COMP/TSP-1 c
32	507	7.7	1068	22 AAG67244	Amino acid sequenc
33	476.5	7.2	877	22 AAU16959	Human novel secret
34	472.5	7.2	800	22 ABB10492	Human cDNA SEQ ID
35	472.5	7.2	800	22 AAU18148	Novel human uterin
36	472.5	7.2	800	22 AAU17031	Human novel secret
37	424	6.4	1584	20 AAM99300	Human BAI1 protein
38	418	6.3	1522	20 AAM99302	Human BAI3 protein
39	418	6.3	1522	21 AAB23601	Human secreted pro
40	415	6.3	1075	22 AAU32291	Novel human secret
41	400	6.1	999	21 AAY94990	Human secreted pro
42	390	5.9	1093	22 AAE02455	Mouse semaphorin G
43	388.5	5.9	810	18 AAM37500	Human nel-related
44	384.5	5.8	1572	20 AAM99301	Human BAI2 protein
45	382	5.8	780	22 AAG67241	Amino acid sequenc

ALIGNMENTS

RESULT 1	
AAB19677	
ID AAB19677 standard; Protein; 1172 AA.	
XX	
AC AAB19677;	
XX	
DT 05-FEB-2001 (first entry)	
XX	
DE Human thrombospondin-2.	
XX	
KW Thrombospondin-2; TSP-2; human; angiogenesis; cell proliferation; melanoma; tumour; cancer; squamous cell carcinoma; antiangiogenic; prostate cancer; psoriasis; rosecea dermatosis; antitumour; therapy.	
XX	
OS Homo sapiens.	
XX	
PH Key	Location/Qualifiers
FT Region	382..429
FT	/note= "type 1 repeat"
FT Region	384..390
FT	/note= "WSPNAEW sequence involved in antiangiogenic activity"
FT	
FT Region	438..490
FT	/note= "type 1 repeat"
FT Region	495..547
FT	/note= "type 1 repeat"
XX	
PN WO200057899-A1.	
XX	
PD 05-OCT-2000.	
XX	
PF 24-MAR-2000; 2000WO-US07835.	
XX	

PR 31-MAR-1999; 99US-0127221.
XX (GEO) GEN HOSPITAL CORP.
XX
PI Detmar M, Streit M;
XX
DR WPI; 2000-656131/63.
DR N-PSDB; AAA88669.
XX
PT Treating a disorder characterized by unwanted cell proliferation e.g. precancerous, cancerous or neoplastic cells or presence of tumor
PT preferably of skin or prostate, comprises increasing thrombospondin-2
PT activity
XX
PS Disclosure; Fig 2; 73pp; English.
XX
CC The present sequence is that of human thrombospondin-2 (TSP-2).
CC The invention is based on the discovery that overexpression of
CC TSP-2 decreases tumor size in vivo, and features methods for
CC modulating unwanted angiogenesis and tumour growth. Treatment of
CC unwanted cell proliferation or angiogenesis involves increasing
CC TSP-2 activity. This is achieved by administering an agent which
CC increases TSP-2 activity, especially a TSP-2 polypeptide, a TSP-2
CC derived polypeptide or retro-inverso peptide, a nucleic acid
CC encoding TSP-2, an agonist of TSP-2, or an agent that increases
CC TSP-2 gene expression. The TSP-2 polypeptide may include at least
CC one type 1 repeat such as the WSPAEW peptide (see AAB19683).
CC which is involved in the angiogenic activity of TSP-2. The
CC method is used to treat a disorder characterised by pre-cancerous,
CC cancerous or neoplastic cells, or the presence of a tumour, or a
CC disorder that affects epithelial tissues resulting in unwanted
CC skin cell proliferation. Such disorders include malignant
CC melanoma, prostate cancer, squamous cell carcinoma, aged skin,
CC rosacea dermatosis, psoriasis, and skin damage caused by
CC photoradiation (all claimed). Evaluating the presence of TSP-2
CC nucleic acid or protein is useful for diagnosing a subject at risk
CC of unwanted cell proliferation or angiogenesis. Methods are also
CC provided of identifying compounds that modulate TSP-2 activity.
XX
SQ Sequence 1172 AA;

Query Match 100.0%; Score 6605; DB 21; Length 1172;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYRLVLLALWVPSQAGHQBKDTFDFLSISNLRKTIGAKQRPDPGPVAYRFRF 60
DB 1 MWRLVLLALWVPSQAGHQBKDTFDFLSISNLRKTIGAKQRPDPGPVAYRFRF 60

QY 61 DYIPVNNADLSKITKIMRQKSGFELTAQLKQDGKSRGTLTLLALEGPGLSQRQFEIVSNGP 120
DB 61 DYIPVNNADLSKITKIMRQKSGFELTAQLKQDGKSRGTLTLLALEGPGLSQRQFEIVSNGP 120

QY 121 ADTLTLTWIDTRHVSLEVDGLADSQWKNVTVOVAGETYSLSHVGCDDLIDSFALDEPFY 180
DB 121 ADTLTLTWIDTRHVSLEVDGLADSQWKNVTVOVAGETYSLSHVGCDDLIDSFALDEPFY 180

QY 181 EHLQAEKSMYVAKGARSHPHRLGVLNHLVFNSEVEDILSKGCGQGGAGAEINAISEN 240
DB 181 EHLQAEKSMYVAKGARSHPHRLGVLNHLVFNSEVEDILSKGCGQGGAGAEINAISEN 240

QY 241 TETRLGLPHVTTEYVGFSSEREPEYCSCEELGNMVELSLHLVNLQPSLENLRKVSND 300
DB 241 TETRLGLPHVTTEYVGFSSEREPEYCSCEELGNMVELSLHLVNLQPSLENLRKVSND 300

QY 301 NOFLHELIGGPKTRNMSACWQDGRFAENETWVVDSTCTCTCKFKTICHOITCPPATC 360
DB 301 NOFLHELIGGPKTRNMSACWQDGRFAENETWVVDSTCTCTCKFKTICHOITCPPATC 360

QY 361 ASPSFVEGECPCSLHSDVDEGFWPAPWTCVTCGSGTQGRSCDVTNTCLGPSI 420
DB 361 ASPSFVEGECPCSLHSDVDEGFWPAPWTCVTCGSGTQGRSCDVTNTCLGPSI 420

QY 421 QTRACSLSKCDTRIRQDGGSHWSPWSSCVTCGVGNITRILCNPSVPQMGKCKGSG 480
DB 421 QTRACSLSKCDTRIRQDGGSHWSPWSSCVTCGVGNITRILCNPSVPQMGKCKGSG 480

QY 481 RETKACQCAPCIDGRWSPWSPWSSACTVTCAGIRIRTRVCNSPEPQYGGKACVGDVQER 540
DB 481 RETKACQCAPCIDGRWSPWSPWSSACTVTCAGIRIRTRVCNSPEPQYGGKACVGDVQER 540

QY 541 QMCNKRSCFVDCCLSNPCFPAGQSCSPDGSWSCSPVGLNGTHCEDLDECALVPDI 600
DB 541 QMCNKRSCFVDCCLSNPCFPAGQSCSPDGSWSCSPVGLNGTHCEDLDECALVPDI 600

QY 601 CFSTSKVPCVNTOPGFHCLPCPPRYRGVQPVGLEAAKTEKQVCEPENCKDKTHNCH 660
DB 601 CFSTSKVPCVNTOPGFHCLPCPPRYRGVQPVGLEAAKTEKQVCEPENCKDKTHNCH 660

QY 661 KHAECIYLGHSFSDPMYKCECOTGYAGDLICGEDSDLGWPNLNLVCATNATYHCIDKNC 720
DB 661 KHAECIYLGHSFSDPMYKCECOTGYAGDLICGEDSDLGWPNLNLVCATNATYHCIDKNC 720

QY 721 PHLNPSGOEDPKDGIQDACCDDDDNDGVTDEKNCOLLFNPRQADYDKDEVGDRCDNCP 780
DB 721 PHLNPSGOEDPKDGIQDACCDDDDNDGVTDEKNCOLLFNPRQADYDKDEVGDRCDNCP 780

QY 781 YVHNPAQIDTNNNGEGDACSVDIDGDDVFNERNDCPYVYNTDQRTDGDGVGDHCDNCPL 840
DB 781 YVHNPAQIDTNNNGEGDACSVDIDGDDVFNERNDCPYVYNTDQRTDGDGVGDHCDNCPL 840

QY 841 VHNPDQTDVNDLVGDCQNNEDIDDDGHQNNQDNCPIYISNANQADHQRDGGDADCDPDD 900
DB 841 VHNPDQTDVNDLVGDCQNNEDIDDDGHQNNQDNCPIYISNANQADHQRDGGDADCDPDD 900

QY 901 DNDGVDPDRDRCLVFNPDQEDLDGDRGDIKDDFDDNDNIPDIDVCPENNAISETDPR 960
DB 901 DNDGVDPDRDRCLVFNPDQEDLDGDRGDIKDDFDDNDNIPDIDVCPENNAISETDPR 960

QY 961 NFQMVPLDPKGTQIDPNWIRHOGKELVQTANSOPGIAVGDFEFGSDVSTFYVNTDR 1020
DB 961 NFQMVPLDPKGTQIDPNWIRHOGKELVQTANSOPGIAVGDFEFGSDVSTFYVNTDR 1020

QY 1021 DDDYAGFVFGVQSSRRFVVMKQVTOYWEQDPTFRAYGSGVSLKVYNSSTGTGHELRN 1080
DB 1021 DDDYAGFVFGVQSSRRFVVMKQVTOYWEQDPTFRAYGSGVSLKVYNSSTGTGHELRN 1080

QY 1081 ALWHTGNTPGVORTLWHPDRNIGWKDYATYRWHLTHRPKTYIRVLVHEGKQVMADSGPI 1140
DB 1081 ALWHTGNTPGVORTLWHPDRNIGWKDYATYRWHLTHRPKTYIRVLVHEGKQVMADSGPI 1140

QY 1141 YDQTYAGGRGLFVFSQEMVYFSDLKYBCRDI 1172
DB 1141 YDQTYAGGRGLFVFSQEMVYFSDLKYBCRDI 1172

RESULT 2
AAB00043
ID AAB00043 standard; Protein; 1172 AA.
XX
XX AAB00043;
XX
XX 08-NOV-2000 (first entry)
XX
XX Human thrombospondin-2 (TSP-2).
KW TSP-1; TSP-2; COMP; cartilage oligomeric matrix protein;
KW thrombospondin; angiogenesis; tumour; treatment; cancer;
KW arthritis; psoriasis; diabetic retinopathy; corneal graft rejection;
KW glaucoma.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Region 381..436

FT	Region	/label= Type 1 repeat region 437..493	301	Db	301	nqflwelligppktrnmsacwqdgdrffaenctwvvdscctctckkfkfichqitcpptac	360
FT	Region	/label= Type 1 repeat region 494..550	QY	361	ASPSFVEGECPCSLHSVDGEGSWPWAEMTQCSTVTCGSGTQQRGSRCDVTSNTCLGPSI	420	
FT	Region	/label= Type 1 repeat region	Db	361	aspsfvegeccpsclhsvdgeegswpwaewtqcsvtcsgtqqrgrscdvtsntclgpsi	420	
XX	WO200044908-A2.		QY	421	QTRACSLSKCDTRIRIQDGGSHSWPSSCSTVCVGNITRIRLNCNSPWPQGGKNCXGSG	480	
PN	03-AUG-2000.		Db	421	qtracslskcdtririqdgshswpsscsvtcgvgnitrlrlcnsppvgmggkncksgs	480	
XX	01-FEB-2000; 2000WO-US02482.		QY	481	RETKACOGAPCIDGRSPWSPWACTVTCAGGIRETRVCNSPEPOVGGKACVGDVGER	540	
XX	01-FEB-1999; 99US-0118053.		Db	481	retkacogapcidgrwspwpsactvtcaggiretrvcnspepvggkacvgdger	540	
XX	(BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.		QY	541	QMCNKRSCPVDGCLSNPCFPGAQCSFPDGSGWSCGSPVGLNGTHCEDLDECALVPDI	600	
PA	Lawler JW;		Db	541	qmcnkrcspvdgclsnpcfpgaqcspfpgsgwscgfpvgflngthcedldecavpdi	600	
PI	WPI; 2000-514823/46.		QY	601	CFSTSKVPRCVNTQPGFHCPLPCPPRYRGNQPVGVLGAATEKQVCPBPENCKDKTHNCH	660	
DR	Nucleic acids encoding chimeric proteins such as cartilage oligomeric		Db	601	cfstskvprcvntqpgfhcplpcppryrgnpgvvgvgleaatekqvcpenpckdkthnch	660	
PT	matrix protein (COMP)/thrombospondins (TSP)-1 and 2, useful for		QY	661	KHABCIYLGHFSDPMYKCEQGTGAGDGLICGEDSDLDGWPENLNLVCAATNATYHCIDNC	720	
PT	Inhibiting angiogenesis and treating diseases such as cancer		Db	661	khabciylghfspdmykceqgtgagdgllcgedsdldgwpnlvcatnatyhcikdnc	720	
XX	Disclosure; Fig 2; 40pp; English.		QY	721	PHLPNSGOEDFDKDIGDACDDDDNDGVTDEKDNCCOLLFNPRQADYDKDSEVGRDCNCP	780	
XX	New nucleic acids are described which encode a protein comprising		Db	721	phlpnsgefdkdgigdacddddndgvtdekdnccollfnprqadydkdevgrdcnncp	780	
CC	the second and third type-1 repeats of human TSP (thrombospondin)-1,		QY	781	YVHNPAQIDTNNEGEGDACSVDIDGDDVFNERNDCPVYNTDQRTDGDGVDGDCNCP	840	
CC	but not the TGF (transforming growth factor)-beta activation region		Db	781	yvhnpaqidtnnegegdacsvdidgddvfnernncpvvntdqrdtdgdvgdghdcncpl	840	
CC	of human TSP-1. The nucleic acid of TSP (thrombospondin)-1 containing		QY	841	VHNPQDQTDVNDLVGDCDNNEDIDDDGHQNNQDNCYPYISNANQADHRRDGGQDACPDD	900	
CC	the second and third type-1 repeats and the COMP (cartilage		Db	841	vhnpgdtdvndlvvgdcddnnediddghqnnqdnccypysnanaqadhrdggdaccpdd	900	
CC	oligomeric matrix protein) assembly sequence (COMP/TSP-1) was		QY	901	DNDGVDPDRNCRLVFNPDDELDGDRGDIKDDFNDNIPDIDDDVCPENNAISETDFR	960	
CC	produced by PCR (polymerase chain reaction). Expression of COMP/TSP-1		Db	901	dndgvdpdrncrlvfnpdqedi dgdrgidckdfndnlpdi ddvcpennaia setdfr	960	
CC	caused inhibition of the growth of tumours in mice models.		QY	961	NFQWVLPDPKGTTQIDPNWIRHOGKELVOTANSDPGIAVGDFSGVDFSGTFYVNTDR	1020	
CC	Thus the nucleic acids and proteins may be useful for treating		Db	961	nfgwvlpdpkgttqidpnwvtrhoggkeltvotansdpgiavgdfsgvdfsgtfyvntdr	1020	
CC	angiogenesis related diseases such as cancer (by reducing the rate of		QY	1021	DDYAGVFGYQSSSFYVVMKQVTQTYWEDQPTRAYGSGVSLKVVNNTGTGTEHLRN	1080	
CC	growth and size of tumours), arthritis, psoriasis, diabetic		Db	1021	ddyagvfgyqsssfyvvmkwqvtqtywedqptraygsgvslkvvnstgtgtehlrn	1080	
CC	retinopathy, corneal graft rejection, and glaucoma. They may also be		QY	1081	ALWHTGNTPGQVRLWHDPRNIGWKDYATYRWHLTHRPKTYIRVLYHEGKQVMADSGPI	1140	
CC	used for treating human immunodeficiency virus (HIV) infection.		Db	1081	alwhtgntpgqvrtlwhdprnigwkdyatyrwhlthrpktgyirvlyhegkqvmdsgpi	1140	
CC	Anti-angiogenic therapy has little toxicity, does not require the		QY	1141	YDQTYAGGRUGLGFVFSQEMVYFSDLKVCECRDI	1172	
CC	therapeutic agent to enter tumour cells or cross the blood-brain		Db	1141	ydttyaggruglgfvsqemvyfslkvecrdi	1172	
CC	barrier, controls tumour growth independently of growth of		QY	1172	YDQTYAGGRUGLGFVFSQEMVYFSDLKVCECRDI	1172	
CC	tumour cell heterogeneity, and does not induce drug resistance.		Db	1172	ydttyaggruglgfvsqemvyfslkvecrdi	1172	
XX	Sequence 1172 AA;		RESULT	3			
QY	Query Match	99.5%; Score 6570; DB 21; Length 1172;	AAB90800				
QY	Best Local Similarity	99.6%; Pred. No. 0;	ID	AAB90800 standard; Protein; 1170 AA.			
QY	Matches 1167; Conservative	0; Mismatches 5; Indels 0; Gaps 0;	XX	AAB90800;			
QY	1 MWRLVLLALWWPSTQAGHODKDTTDLFSISNINRKTIGAKOFGGPDGPVAYRVR	60	XX	15-JUN-2001 (first entry)			
Db	1 mwrlvllalwwpstqaghdkttdtldfslsinnrktigakqfgrpdgpvayrvrf	60	XX	Human shear stress-response protein SEQ ID NO: 100.			
QY	61 DYIPPNVADDLKITKIMRQEGFFLTAKLKQDGKSGTLALLEGCLSORQEIFVNSGP	120	XX	Human; shear stress-response protein; vascular disease;			
Db	61 dyppnvaddlkitkkmrqegffltaklkqdgksrgtlallegpqlsqrfvnsngp	120	XX	arteriosclerosis.			
QY	121 ADTLDLTWIDGTRHVVSLDEVDGLADSQWKNVTQVAGETYSLSHVGCDLDSFALDPFY	180	OS	Homo sapiens.			
Db	121 adtldltwidgehrhvvsldevdgladsqwknvtqvagetyslhvgcdllgpvaldepfy	180					
QY	181 EHLQAEKSRMYVAKGSAHSHFRLGLQNLVHLVFNSEVEDLSKKGCGGQGAELNAISEN	240					
Db	181 ehlaeksrmyvakgsahshfgrllqnlvhlvfnsevedlskkgcgqgggaelnaisen	240					
QY	241 TETLRGLPHVTVTVGSSRRRPPVCRSCCEELGNMVQELSGHLVNLQPSNLKRVSN	300					
Db	241 tetrlrglphvttvvgssrrppvcrscceelgnmvqelsghlvnlqpsnlkrvsnd	300					
QY	301 NQFLWELIGGPPKTRNMSACWQDGDRFFAENCTWVVDSCCTCTCKKFKTICHQITC	360					
Db	301 nqflwelligppktrnmsacwqdgdrffaenctwvvdscctctckkfkfichqitcpptac	360					

XX WO200125427-A1.
 XX 12-APR-2001.
 XX 02-OCT-2000; 2000WO-JP06840.
 XX 01-OCT-1999; 99JP-0280976.
 XX (KYOW) KYOWA HAKKO KOGYO KK.
 PA (NOJII) NOJIMA H.
 XX Nojima H, Yoshisue H, Obayashi M, Ota T, Kawabata A, Sakurada K;
 PI Kuga T, Sekine S, Nakamura Y, Sugano S;
 XX WPI: 2001-266308/27.
 DR N-PSDB: AAH02923.
 XX
 XX DNA sequences, proteins encoded by them and antibodies against them
 PT useful in diagnosis and treatment of vascular disease caused by
 PT arteriosclerosis
 XX
 PS Claim 60; Page 515-521; 678pp; Japanese.
 XX
 CC The present invention provides the protein and coding sequences of a
 CC number of human shear stress response proteins. These are useful in the
 CC diagnosis, treatment and screening of vascular diseases caused by
 CC arteriosclerosis, including heart failure, post-PTCA restenosis and
 CC hypertension.
 XX
 SQ Sequence 1170 AA;
 6;
 Query Match 64.0%; Score 4228; DB 22; Length 1170;
 Best Local Similarity 61.6%; Pred. No. 6.6e-253;
 Matches 725; Conservative 170; Mismatches 267; Indels 14; Gaps
 QY 1 MWRL-VLLALWVWPSTQAGHQDKD-TTFDLFSISNINRKTIGAKQFRGPPGVPAYRFV 58
 DB 3 lawglvflmhvcotnrupesgdnsvdfelgaarksgrrlvkgpdpapafie 62
 QY 59 RFDYIPVNNADLSKITKIMQKEGFFLTAQKDGKRGTLLEGLSGLVHVNQPSNLK 118
 DB 63 danlppvddkfqlvdavraekflllaslrqmktrgtllalerkdhsqgvfsvsn 122
 QY 119 GPADTLDITYWIDTRHVVLEVDGLADSONKNTVQVAGETYSLVHVCDDLIDSPALDEP 178
 DB 123 kgagtlidlsiltvgkghvsvveaalatgqwkstifvgedraqlyidckmenaeldvp 182
 QY 179 F---YEHLQAEKSRMYAKGSAHSHFRGLLQNVHLVPENSVEDILSKKCGQOGAETN 235
 DB 183 lqsvftrdlaslarlriakgvnd-nfgvlgqvrvfvtgtpedllrnkgcsstsvilt 241
 QY 236 AISENTEFLRLGPHVTVTYGVSSRRPEVCSERCEELGNMVQELSGHLVHVNQPSNLK 295
 DB 242 L--dnvngvsgpairstnyghktkdqalcgiscdelssmvlrgritvltlqdsir 299
 QY 296 RYSDNQFLWELIGPPKTRNNSACWDGRFAENETWVDSCTCTCKKFTICHOJTC 355
 DB 300 kvtcenkelanelirrp-----lcyhngvqyrnneetvdsctechcqnsvtlckkvc 353
 QY 356 PPATCASPSFVGECCFSLHSDVEEGSWPAEWTCQSVTCGSGTQQRGRSCDVTNCT 415
 DB 354 plmpcsnatvpgcepcrcwpsdsaddgswpsewtscstscnglqgrfscdlnncr 413
 QY 416 LGPSITQTRACLSKCDTRIQDGGHSHSPWSSCSVTCGVGNITRILCNPSVPOMGKN 475
 DB 414 egssvqtrtchicqckrfkqggghswspsscsvtcgvgvtrilcnpspqqmgkp 473
 QY 476 CKGSGRETKACOGAPCPIDGRSWPSWSACTVTCAGGIRETRVCNSPEQYGGKACVG 535
 DB 474 cegaretkackckacpningvpswpsdlcstvcggvqvrksrlnncnpapqfggkdcvg 533

QY 536 DVOERQMCNKRSCPVDCGLSNPCRPAGOCSSFPDGSWCGSCPVGFLNGTHCEDLDECA 595
 DB 534 dvtengicnkdcpidgclsnpcrfagvictaypdsqwkcgacppysnglqctavdeck 593
 QY 596 LVPDIFCSTSKVPCVNTQPGFHLCLPCPPRYRGNOVPVGVLEAAKTEKOVCEPENCKDK 655
 DB 594 evpdacfnhgehrcentdpynclpcprftgsqpfggvehatankqvckprnpctdg 653
 QY 656 THNCHKHAEICYLGHFSDPMYKCECOTGYAGDGLICGSDSLDGPNNLNCATNATVHC 715
 DB 654 thdcnkakcnlyghysdpmrceckpyagngilicgedtdldgwpnenlvcvanatync 713
 QY 716 IKDNCPLHNSQEDFDKDGIGDADDDDDNDGVVDERDNCOLLFNPRQADYDKDEVGDR 775
 DB 714 kkdncplnpsagedykdigdadcdndndkdpdrncpfnypaqdydrddvgr 773
 QY 776 CQNCPIVHNPAQIDTDNNGEGDACSVDIDGDDVFNERONCPIVYNTDQDRTDGDGVGDH 835
 DB 774 cdncpnhnpqadtndnngedacaadldgdilnerdncgyvynvgrtdmdvgdgc 833
 QY 836 DNCPLVHNPQDQDNDLVGQCDNNEDIDDDGHONNODNCPYISNANOADHRRDQGDGA 895
 DB 834 dncplehnpdqldsdridgtcdnnqgldedgngnldncpyvnanqadhdkgkda 893
 QY 896 CDPDDNDGVDDRNCRLVFNPDQEDLDGDRGCDICDDFDNDNIPDIDDVCPENNAIS 955
 DB 894 cdhddndgipdkdncrlvpnpdqkdsdgdrgrackddfdhdsvpdiddicpenvids 953
 QY 956 ETDPRNFQWPLDPKGTQIDPNWVIRHQKELVQANSDPGIAVGFDEFGSVDSGTFY 1015
 DB 954 etdfrfqpildpkgtsgndpnwvrrhggkelvtvncdpglavgydefnavdfsgtff 1013
 QY 1016 VNTDRDDYAGFVGYQSSRFYVVMKQVTVYWEQDPTRAYGSGVSLKVVNSTTGTG 1075
 DB 1014 interddayagfvfygqssrfyvmwkvqtsywdntnptraqgysglsvkvnsttggp 1073
 QY 1076 EHLRNALWHTGNTPGQVRLMHPDPRNIGWKDYATAYRWLTHRPKTYIRVLVHEGQVMA 1135
 DB 1074 ehlnalwhtgntpgqvrtlwhdprhgwkdftayrwlshrpkgtfgrvrmvgegkima 1133
 QY 1136 DSGPIYDQTYAGRGLGLEVFQEMVYFSDLKVECRD 1171
 DB 1134 dsdpiydktyaggrlglvfseqemvfdsdkyecd 1169
 RESULT 4
 AAB74450
 ID AAB74450 standard; Protein; 1170 AA.
 XX AC AAB74450;
 XX DT 06-JUN-2001 (first entry)
 XX DE Human variant thrombospondin 1.
 XX KW Human; variant thrombospondin 1; variant thrombospondin 4; SNP;
 KW polymorphism; vascular disease; coronary artery disease; forensics;
 KW myocardial infarction; atherosclerosis; stroke; venous thromboembolism;
 KW pulmonary embolism; paternity test.
 XX OS Homo sapiens.
 XX PN WO200118250-A2.
 XX PD 15-MAR-2001.
 XX PF 07-SEP-2000; 2000WO-US24503.
 XX PR 10-SEP-1999; 99US-0153357.
 PR 26-JUL-2000; 2000US-0220947.
 PR 16-AUG-2000; 2000US-0225724.
 XX (WHED) WHITEHEAD INST BIOMEDICAL RES. FA

(MILL-) MILLENNIUM PHARM INC.
Lander ES, Gargill M, Ireland JS, Bolik S, Daley GO, McCarthy JJ;
WPI: 2001-226749/23.
N-PSDB; AAF95238.
Nucleic acids comprising single nucleotide polymorphisms, useful in applications such as forensics, paternity testing, medicine, genetic analysis and phenotype correlations to diseases such as diabetes and atherosclerosis -
Claim 20; Fig 1; 242pp; English.
The present invention provides a method of diagnosing a vascular disease in an individual, involving determining the sequence at various polymorphic sites within the human thrombospondin 1 and thrombospondin 4 genes. The sequences at a number of polymorphic sites are also provided in the specification. In particular, the method can be used in the diagnosis of atherosclerosis, myocardial infarction, coronary heart disease, stroke, peripheral vascular diseases, venous thromboembolism and pulmonary embolism. Single nucleotide polymorphisms (SNPs) are also useful in forensics, paternity testing, genetic analysis and phenotype correlations to diseases. The present sequence is the human variant thrombospondin 1 protein.
Sequence 1170 AA;
Query Match 54.0%; Score 4228; DB 22; Length 1170;
Best Local Similarity 51.6%; Pred. No. 6.6e-253; Indels 14; Gaps 6;
Matches 725; Conservative 170; Mismatches 267;
QY 1 MWRL-VLLALNVPSTQAGHDKD-TTFDLFSISNINKTKTAKQFGRPDGVPAYRFV 58
Db 3 lawglvflmhcvtgnripesgdnsvdfelgtgaarkgsgrllvkgpdpapafrie 62
QY 59 RFDYTPPVNADLSKITMRQKEFFLTAQLKQDGKSGRTLLALEGGLSORQFEIVSN 118
Db 63 danllppvddkfdlvdavraekflllaslrqmkktgrtllalerkdhsqgvsvvsn 122
QY 119 GPADTLTYIDGTRHYVLESDVGLADSONKNTVAVAGETSLHVGCDLIDSLFALDEP 178
Db 123 gkagtlidslvtgqkqhvsveallatqgwkstlfvqdradqlvdeekmenaelvdp 182
QY 179 F---YEHQAERMYVAKGSAESHFRGLLQNVHLVFNSEVEDILSKKGGQGGGAETN 235
Db 183 lqsvftrdlaslarlriakgvnd-nfgvqlqnvrfvgttpealrnkgsstsvilt 241
QY 236 AISENTEILRLGPHVTEYVGSRRPVEVCSEELGNMVOELSLHLVLPQSENLK 295
Db 242 l--dnvnvngspairtnyghktdlqalcglsdelssmvlairgtivttlqdsir 299
QY 296 RVSNDNQFLWELIGPPKTRNNSACWQDGRFAENETWVVDSCCTCTCKKFTICHQITC 355
Db 300 kvtreenkelanelrtp-----lcyhngvqyrneetvdsctechcqnsvtckkvc 353
QY 356 PPATCASPFVBECEPCSLHSDVEEGSWPAEWTCQSVTCGRTQGRSCDVTNFC 415
Db 354 plmpcsnatvpdgeccprcwpsadsaddgswpsewtscstcnglqgrgrscdslnnrc 413
QY 416 LGPSTQTRACSLSKDTRTRQDGGSHSWSPSSCSVTCGVGNITRLCNLSPVPMQGNK 475
Db 414 egssvqtrtchigecdkrfkqdgghswspsscsvctcgvgvtriricnspsgmqngk 473
QY 476 CKSGRETKACGACGPFIDGRISPNPSACTVTCAGGIRERTRVCNSPEPQYGGKACV 535
Db 474 cegaretkackdcpingvgwpswdicvctcggvgqkrslcnnpapqfgkdcvg 533
QY 536 DYQERQMKRSCVPDVCGLSNPCFFPGAQCSFPDGSWSGCGSPVGLNGTHCEDLDECA 595
Db 534 dvtenqicnkqcdpdcglcslncpfcagvkctsydpdgsdkcagppysnglqctdvdeck 593

QY 596 LVPDICFSTSKVPRCVNTOPGFHCLPCPPRYRGNOVPVGLEAAKTEKQVCSBPENCKOK 655
Db 594 evpdacfnhnghrceidtdpgvncipcpbrftgspqfggvehatankqvcprnptcdg 653
QY 656 THNCHKHAEIYLGHFSDPMYKCEQOTGYAGDGLICGEDSLDGGWNLNLVCATNATYHC 715
Db 654 thdcnkakcnlyghysdpmrceekpgyagngilcgedtdlgwpenlvcvanatyhc 713
QY 716 IKDNCPLHNSQGEDEFKDGIGDADDDDDNDGVTFDEKNCQLLENPROADYDKDEVGDR 775
Db 714 kdkncpnlpnsgeydkdgidacddndkdpddrdncpfhynpacydydrdvgar 773
QY 776 CQNCPIVHNPAQIDTNDNGEGDACSVDIDGDDVFNERDNCPIYVNTDQRTDGDGVDGHC 835
Db 774 cncpynhnpdadtndngegacaadidgdgillnerdncgyvynvqgrtdmdvgdgc 833
QY 836 DNCPLVHNPQDQTDVNDLVGDDCDDNEDIDDDGHQNNQNCPIYSANQADHRRDQGDGA 895
Db 834 dncpllehnppdqldsdgsdrigtdcddmqldedghqnnlndcypvnpnagadhdkgda 893
QY 896 CDPPDDNDGVPDDRNCRLVFNPDQEDLDGDRGDRICKDDFDNDNIPDIDDDVCPENNAIS 955
Db 894 cdhddndgipdkdncrlvnpdpqkdsdgdrgdackdhdsvpdlddicpenvdis 953
QY 956 ETDFFNFQWPLDPKGTQTQIDPNWIRHQCKELVQTANSDPGIANVGFDFSGVDFSGTFY 1015
Db 954 etdfrfmgipdkgtsgndpnnvvrhggkelvqtnvncdpqlavgydefnavdsgtff 1013
QY 1016 VNTDQDDYAGFVGYQSSRFYVVMKQVOTQTYWEDQPTRAYSGVSLKVVNSTGTG 1075
Db 1014 interddayagfvfygssrfyvmvwmkvtqsywdntptraqgysglsvkvnstgpg 1073
QY 1076 EHLRNALMHTNTPGQVFTLWHDPRNIGWKDYATAYRHLTHRPKTYGIRLVHVEGQVMA 1135
Db 1074 ehlrnalwhtnpgqvtlwhdprhgwkdftayrwlshrpktgfrvnmvegkima 1133
QY 1136 DSGPIYDQTYAGRGLGFVFSQEMVYFSDLKVECRD 1171
Db 1134 degpiydktyagrglgrfvsqemvifsdlykcecd 1169
RESULT 5
AAB00042
ID AAB00042 standard; Protein; 1152 AA.
AC AAB00042;
XX
DT 08-NOV-2000 (first entry)
XX Human thrombospondin-1 (TSP-1).
DE
XX TSP-1; TSP-2; COMP; cartilage oligomeric matrix protein;
KW thrombospondin; angiogenesis; tumour; treatment; cancer;
KW arthritis; psoriasis; diabetic retinopathy; corneal graft rejection;
XX glaucoma.
OS Homo sapiens.
XX Key Location/Qualifiers
FH Region 361..416
FT /label= Type 1 repeat region
FT Region 417..473
FT /label= Type 1 repeat region
FT Region 474..530
FT /label= Type 1 repeat region
XX WO2000044908-A2.
XX
XX 03-AUG-2000.
XX
XX 01-FEB-2000; 2000MO-US02482.
XX
XX 01-FEB-1999; 99US-0118053.
PR

XX WPI: 2001-235357/24.
DR N-PSDB; AAH34535.
XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS Claim 11: Page 7409-7411; 9803pp; English.
XX
CC AAH32943 to AAH37195 and AAH37514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patient's own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps;
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAG77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ Sequence 571 AA;

Query Match 49.2%; Score 3251; DB 22; Length 571;
Best Local Similarity 100.0%; Pred. No. 6.8e-193;
Matches 568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 605 SKVPRCVNTQPGFHCIPCPRRYGNQPVGVGLEAAKTEKQVCEPENPCPKDKTHNCHKHAE 664
DB 4 skvprcvntqpgfchelpcrrygnqpvvgvgleaaktekqvcpenpcpkdkthnchkhae 63
QY 665 CYYLGHFSDPMKCEQCTQYAGDGLICGEDSLDGNPNLVCATNATYHCIKDNCPLHP 724
DB 64 cyylgfhfsdpmkcectqyagdgllcgdsldgwpnlvcatnatyhcikdncplhp 123
QY 725 NSGOEDFKDGTGACDDDDNDGVTDEKDNLCQLLENPRQADYDKEVGDRCDCNCPYVHN 784
DB 124 nsqgedfkdgtdgacddddndgvtdekdncqlfnprrqadydkdevgdrcdncpyvhn 183
QY 785 PAQIDTDNNGEGDACSVDIDGDDVFNERDNCPPVYNTDQDRTDGGVGDHDCNCPVHNHP 844
DB 184 paqidtdnngedacsvdldgddvfnerncnpvyntdqrdtdgvgdhdncpvhnp 243
QY 845 DQFDVNDLVGQCDNNEIDDDGHQNNQDNCPIYSNANQADHRDQGDGACDPPDDNDG 904
DB 244 dqfdvndlvqcdnnediddghqnqndncpiysnanqadhrdrggdcacpddndg 303
QY 905 VPDRDNCRLVNPQDELDGDRGDIKDDFDNDNIPIIDVCPENNAISDTERFOM 964
DB 304 vpdrdncrlvnpdqdeldgdrgdickddfdndnippiidvcpennaisdterfom 363
QY 965 VPLDPPGTTQIDPNMWRHOGKELVOTANSDPGIAVGDFGFSVDFSGFYFNTDRDDY 1024
DB 364 vpldpkgttqidpnwvrhlgkqlvotansdpgiavgdfgsvdfsgtfyvtndrddy 423
QY 1025 AGVFYQYSSSRFVVMWQVQTYWEDQPTRAYGVSGLKVVNSSTTGTGHELRNALWH 1084
DB 424 agvfyyqssrfvvmwqvtqywedqptraygysglkvvnsttgtghehrnalwh 483
QY 1085 TGNTPQCVFTLWHDPRNIGKDYATYRWHLTHRPKTYIRLVVHEGKQVMADSGPIYDQT 1144
DB 484 tgntpqcvftlwhdprnigkdyatyrwhlthrpktgyirvlvhegkqvmadsgpiydt 543
QY 1145 YAGGRGLGVFSEMYVFSDLKVECRDI 1172

DB 544 YAGGRGLGVFSEMYVFSDLKVECRDI 571
RESULT 7
AAB43586
ID AAB43586 standard; Protein; 571 AA.
XX
AC AAB43586;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human cancer associated protein sequence SEQ ID NO:1031.
XX
KW Human; cancer associated gene; cancer antigen; detection; cancer;
KW diagnosis; cytostatic; proliferative; vulnerable; immunomodulator;
KW antidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiviral;
KW antinflammatory; antithyroid; antiallergic; antibacterial; cardiac;
KW dermatological; neuroprotective; thrombolytic; coagulant; nocotropic;
KW vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
KW immune disorder; haematopoietic cell disease; autoimmune disorder;
KW allergic reaction; graft versus host disease; organ rejection;
KW haemostatic; thrombolytic; cardiovascular disorder; infection;
KW neurological disease; drug screening.
XX
OS Homo sapiens.
XX
PN WO200055350-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05882.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI: 2000-587533/55.
XX
DR N-PSDB; AAC77795.
XX
PT Novel isolated nucleic acids comprising sequences encoding peptides
PT useful for treating or diagnosing e.g. cancer -
XX
PS Claim 11: Page 1616-1619; 2352pp; English.
XX
CC AAC77607 to AAC78448 encode the human cancer associated proteins given
CC in AAB43398 to AAB44239. The proteins can have activities based on the
CC tissues and cells the genes are expressed in. Example of activities
CC include: cytostatic; proliferative; vulnerable; immunomodulator;
CC antidiabetic; antiasthmatic; antirheumatic; antiarthritic;
CC antinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
CC dermatological; neuroprotective; cardiac; thrombolytic; coagulant;
CC nocotropic; vasotropic; antipsoriatic and antiangiogenic. The
CC polynucleotides and polypeptides can be used for preventing, treating or
CC ameliorating medical conditions and diagnosing pathological conditions.
CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from
CC the present invention may be used to treat immune disorders by activating
CC or inhibiting the proliferation, differentiation or mobilisation of
CC immune cells, to treat disorders of haematopoietic cells, autoimmune
CC disorders, allergic reactions, graft versus host disease and organ
CC rejection, modulate haemostatic or thrombolytic activity, modulate
CC inflammation, cancers, cardiovascular disorders, neurological disease and
CC bacterial or viral infections. The peptides, nucleotides, antibodies,
CC agonists and antagonists may be also be used in drug screens. AAC78449 to
CC AAC78457 and AAB44240 represent sequences used in the exemplification of
CC the present invention.
XX
SQ Sequence 571 AA;

Query Match 49.0%; Score 3236; DB 21; Length 571;

Best Local Similarity 99.5%; Pred. NO. 5.7e-192;		Matches 565; Conservative 0; Mismatches 3; Indels 0; Gaps 0;	
QY	605 SKVPCVNTQPGFHCILPCPPRYRGNOVGVGLEAAKTERKQVCEPENPCPKDKTHNCHKHAE 664		
Db	4 skvpcvntqpgfchilpcppryrgnpgvgleaaktekxcepncpkdkthnchkhae 63		
QY	665 CIYLGHFSDPMYKCEQOTGAGDGLICGEDSDLDGHPNINLVCAATNATYHCIKDNCPLHP 724		
Db	64 cilyghfsdpmkykceqotgagdglicgedsdlldgwpnlnlvcatnatyhcikdncplhp 123		
QY	725 NSQOEDFDKDGIGDADDDDDNDVTDKDNKQQLFNPRQADYDKDEVGDRCDNCPYVHN 784		
Db	124 nsqoedfdkdgigdadddndvtdkdnkqqlfnprqadydkdevgdrdcncpyvhn 183		
QY	785 PAQIDFNNNGEGDACSVDIDGDVFNERNDCNPYYNTDQDGDGVDGHCNCPLVHNP 844		
Db	184 paqidfnnngedgacsvdidgdvfnerncnpyyntdqdgdgvdgghcncplvhn 243		
QY	845 DQTDVDNLDVGQCONNEDIDGQHNQNDNCPYISNANQADHRRDGGDGDGDDDDNDG 904		
Db	244 dqtdvdnldvgqconnedidgqhnqndncpyisnanaqadhrdggdgdgddndg 303		
QY	905 VPDDRNCRLVFNPOEDLDGSGRGDICKDDFDNDNIPDIDVCPENNAISSETDFRNFOM 964		
Db	304 vpdtrncrlvfnpqedldgsgrgdickddfdndnipdidvcpennaisetdfnfm 363		
QY	965 VPLDPKGTQIDPNVIRHOGKELVOTANSDPGIAVDFDFGVSDFGFFVYNTDRDDY 1024		
Db	364 vpldpkgtqidpnvirhogkelvotansdpgiavdfdfgvsdfgffvntdrddy 423		
QY	1025 AGFVFGYQSSRFYVVMKQVNTQVWEDQPTRAYGSGVSLKVVNSTGCTGHLNALWH 1084		
Db	424 agfvfyqssrfyvwmkqvntqvwwedqptraysgvslkvvnstgctghlnalwh 483		
QY	1085 TGNTPQVRLTWHDPNIGWKDYTYRWHLTRPRTGYIRLVHLEGKQVMDSGPIYDOT 1144		
Db	484 xgntpgvrtlwhdpnigwkdytyrwhltrpkyirvlvhegkqvmdsgpiydot 543		
QY	1145 YAGGRGLFLVFSQEMVYFSDLKYECRDI 1172		
Db	544 yaggrlglfvsqemvyfssdkyecdri 571		
RESULT 8			
AAU02913	AAU02913 standard; Protein; 731 AA.		
XX	AC AAU02913;		
XX	12-SEP-2001 (first entry)		
DE	Angiotensin converting enzyme (ACEV) splice variant, protein #13.		
XX	Angiotensin converting enzyme splice variant; ACEV; interleukin 6;		
KW	granulocyte colony stimulating factor receptor; glucagon; hypertrophy;		
KW	platelet-derived endothelial cell growth factor; cardiovascular disease;		
KW	cellular tumour antigen P53; cyclin-dependent kinase inhibitor 1C;		
KW	vasoactive intestinal polypeptide receptor 2; arteriosclerosis; cancer;		
KW	myocardial infarction; coronary arterial thrombosis; renal disease;		
KW	diabetic nephropathy; muscular disease; immune disorder; sarcoidosis;		
KW	multiple sclerosis; immune complex nephritis; deep vein thrombosis;		
KW	nonaroidotic pulmonary granulomatous disease; endothelial abnormality;		
XX	vascular disorder; asbestosis.		
OS	Homo sapiens.		
XX	WO200136632-A2.		
PN	25-MAY-2001.		
PD	17-NOV-2000; 2000WO-IL00766.		
XX			
FF			
XX			

PR	17-NOV-1999; 99IL-0132978.		
PR	10-DEC-1999; 99IL-0133455.		
XX	(COMP-) COMPUEN LTD.		
PA	Levine Z, David A, Azar I, Khosravi R, Bernstein J;		
PI	WPI; 2001-336004/35.		
XX	N-PSDB; AAS06013.		
DR	Novel alternative splicing variants e.g. variant of angiotensin		
PT	converting enzyme (ACEV), useful in identifying candidate compounds		
PT	capable of binding to the variant and to detect anti-variant antibodies		
XX	Claim 4; Fig 13; 519pp; English.		
XX	The sequence represents an angiotensin converting enzyme splice variant		
CC	(ACEV) polypeptide. The polypeptides of the invention include variants of		
CC	granulocyte colony stimulating factor receptor, glucagon, interleukin 6,		
CC	platelet-derived endothelial cell growth factor, cyclin-dependent kinase		
CC	inhibitor 1C, cellular tumour antigen P53, and vasoactive intestinal		
CC	polypeptide receptor 2. The polypeptides and their associated nucleic		
CC	acids are useful for identification of variant sequences and detection of		
CC	candidate compounds capable of binding to the molecules. The sequences of		
CC	the invention can be used in the treatment and diagnosis of various		
CC	disorders including cardiovascular diseases such as arteriosclerosis,		
CC	myocardial infarction and coronary arterial thrombosis, renal diseases		
CC	such as diabetic nephropathy, muscular diseases such as hypertrophy,		
CC	immune disorders such as immune complex nephritis, multiple sclerosis,		
CC	cancer, sarcoidosis, nonaroidotic pulmonary granulomatous diseases such		
CC	as asbestosis and vascular pathologies involving an endothelial		
CC	abnormality such as deep vein thrombosis.		
XX	Sequence 731 AA;		
QY	Query Match 31.7%; Score 2094; DB 22; Length 731;		
Db	Best Local Similarity 50.1%; Pred. No. 3e-121;		
Db	Matches 365; Conservative 123; Mismatches 227; Indels 14; Gaps 6;		
QY	1 MWRL-VLLALWMPSTQAGHQDKD-TTFDLFSISINRKTIGAKQFQDPDGPVAYRFV 58		
Db	3 lawglvflmhvcegntripesggnsvdfidelfelgaarksgrrlvkqdpssapafrie 62		
QY	59 RFDYIPPNADDLSKITIMRQKEGFFTAQLKQDGKSGTLLALEGPGLSQRQFEIVSN 118		
Db	63 danlippyddkfgdlvdavraekgflillaslrqmktrgtllalerkdhsqgvsvsn 122		
QY	119 GPADTLDLTYYWIDGPRHVVSLEDDVGLADSONKNTVQVAGETYSYSLHVGCDLDSFALDEP 178		
Db	123 gkagcidslsltvqgkqhvsveeallatgqwkstifvqdaqrlgicckmensaelvdp 182		
QY	179 F---YEHQLAERKSMYVARGSARESHFRGLLQNVHLVFNSEVEDILSKKGGQGGAEIN 235		
Db	183 lqsvtrldlaslrlriakgvnd-nfgvlgvqnrvfvtgtpedilrnkgsstsvilt 241		
QY	236 AISENTEFLRGLPHTTYFVGPSSRRPEVCSERCEELGNVWQELSGHLVNLQNPSENLK 295		
Db	242 l--dnvvngvsgspairtnyghktkdlaicgicdelssmvlrigrirtvtlqda 299		
QY	296 RVSDNQFLWELIGPPKTRNMSACWODGRFFAENETVWVDSCTTCTCKKFKTICHQITC 355		
Db	300 kvtreenkelanelrrpp-----lcyhngvqyrnneetvdsctechqnsvtickkvsc 353		
QY	356 PPATCASPFVEGECPCSLHSVDGEGSWPAEWTCQSVTCGSGTQGRSCDVTNVC 415		
Db	354 pmpcsnatvpdgeccprcwpsdsaddgswpsewtscstcgngiqgrgscdlnnrc 413		
QY	416 LPSIQTRACSLSKCDTRIRODGGWSHSPWSSSVTCGVCNITRILCNLPVQMKGK 475		
Db	414 egssvqtrtchicgckrfkqdgqgshwspwsscsvtcgdgvitrlcnspqmgk 473		

QY 476 CKSGRETKACGAPCIDGRSPWSPWSACTVTCAGIRERTRVCNSPEPOYGGKACVG 535
 Db 474 cegaeatckackdaapnggwgpgwspwdicvccggvgkrrlcnmpcpfggkdcvg 533
 QY 536 DVOERQCMKRRCPVDGCLSNPCFPGAQCSFDPGWSGSCVPVGLNGTHCEDLDECA 595
 Db 534 dvtenglcnkqdcpidgclsnpcsfagvcktsypdgswkcgacppgysngiqctdvdeck 593
 QY 596 LVPDIFCSKVPKRCVNTQGFHCLCPPRYRGNQPVGVGLEAAKTEKQVCEPENCKDK 655
 Db 594 evpdacfnhnghearcndpgncldpcprftgsgfvggvehatankgvckprnctdg 653
 QY 656 THNCHKHAECIYLGHPSDPMYKCECTGAGDGLICGEDSLDGLWPNLNLVCATNATYHC 715
 Db 654 thcnknaknynlghysdpmyrceckpgyagngilicgedtldgwpnvlvcvanatyhc 713
 QY 716 IKDNCPLP 724
 Db 714 kkdncplp 722

RESULT 9
 AAB00044
 ID AAB00044 standard; Protein; 757 AA.
 AC AAB00044;
 XX
 DT 08-NOV-2000 (first entry)
 DE Human cartilage oligomeric matrix protein (COMP).
 KW TSP-1; TSP-2; COMP; Cartilage oligomeric matrix protein;
 KW thrombospondin; angiogenesis; tumour; treatment; cancer;
 KW arthritis; psoriasis; diabetic retinopathy; corneal graft rejection;
 KW glaucoma.
 XX
 OS Homo sapiens.
 FH
 FT Key
 FT Region 89..128 Location/Qualifiers
 FT /label= Type 2 repeat region
 FT Region 129..181
 FT /label= Type 2 repeat region
 FT Region 182..226
 FT /label= Type 2 repeat region
 FT Region 227..268
 FT /label= Type 2 repeat region
 XX
 WO200044908-A2.
 XX
 PN 03-AUG-2000.
 PD
 XX
 PF 01-FEB-2000; 2000WO-US02482.
 PR
 XX 01-FEB-1999; 99US-0118053.
 XX
 PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
 XX
 PI Lawler JW;
 XX
 DR WPI; 2000-514823/46.
 XX
 XX Nucleic acids encoding chimeric proteins such as cartilage oligomeric
 PT matrix protein (COMP)/thrombospondins (TSP)-1 and 2, useful for
 PT inhibiting angiogenesis and treating diseases such as cancer
 PS
 XX Disclosure; Fig 3; 40pp; English.
 XX
 CC New nucleic acids are described which encode a protein comprising
 CC the second and third type 1 repeats of human TSP (thrombospondin)-1,
 CC but not the TGF (transforming growth factor)-beta activation region
 CC of human TSP-1. The nucleic acid of TSP (thrombospondin)-1 containing
 CC the second and third type-1 repeats and the COMP (cartilage

CC oligomeric matrix protein) assembly sequence (COMP/TSP-1) was
 CC produced by PCR (polymerase chain reaction). Expression of COMP/TSP-1
 CC caused inhibition of the growth of tumours in mice models.
 CC Thus the nucleic acids and proteins may be useful for treating
 CC angiogenesis related diseases such as cancer (by reducing the rate of
 CC growth and size of tumours), arthritis, psoriasis, diabetic
 CC retinopathy, corneal graft rejection, and glaucoma. They may also be
 CC used for treating human immunodeficiency virus (HIV) infection.
 CC Anti-angiogenic therapy has little toxicity, does not require the
 CC therapeutic agent to enter tumour cells or cross the blood-brain
 CC barrier, controls tumour growth independently of growth of
 CC tumour cell heterogeneity, and does not induce drug resistance.
 XX
 SQ Sequence 757 AA;

Query Match 30.1%; Score 1990; DB 21; Length 757;
 Best Local Similarity 54.1%; Pred. No. 8.3e-115;
 Matches 362; Conservative 72; Mismatches 177; Indels 58; Gaps 6;

QY 549 PVDGCLSNPCFPGAQCSFDPGWSGSCVPVGLNGTHCEDLDECAVDPDIFCSKVP 608
 Db 87 pllhcapgfcfpgvaciqtesgg-rccgpcpagftgngshctdvnecahp--cfpr---v 140
 QY 609 RCVNTQGFHCLCPPRYRGNQPVGVGLEAAKTEKQVCEPENCKDKTHN----- 658
 Db 141 rcntspgfrceacppgysgpthgvgglafakankgvctdinecetgghncvpnsvcint 200
 QY 659 -----CHKHAECIYLGHPSDPMYKCECOT 682
 Db 201 rgfqcgcpcpgfvgdgagcgrgaqfcpdgspsechehadcvl---erdgsrscrv 257
 QY 683 GYAGDGLICGEDSLDGLWPNLNLVCATNATYHCIKNCPLHNSGOEDFDKIGDACDD 742
 Db 258 gwagnilcgrrtdldgfpdekrlrcep---qcrkncvtvpsngedvdrgdgadcdp 314
 QY 743 DDDNDGVTDEKNCQLLFNPRQADYDKVEGDRCDNCPPYVHNPAQIDTNNNGSDACSV 802
 Db 315 daagdyvnekdncplrvnpdqrdtdedkwdacdnrcsqknddktdgtdggrgacdd 374
 QY 803 IDGDDVFNERNDCPPYVNTDQRTDGGVGDHCDNCPVHNPDQTDVNDLVGQDCDNE 862
 Db 375 idgdrirnadncprvpnsdqsdgdlgacdnccpksnpdqadvdhdfvgdacsdg 434
 QY 863 DDDDGCHQNNQDNCPISNANQADHRDGGDGDACDDDDNDGVPDORDNCRLVNPDQED 922
 Db 435 dgdgdghqdsrdncptvpsnagedshdggdcdcdndngvpsdrdnrcrlvnpnpged 494
 QY 923 LDGDRGDIKDDFDNDNIPDIDVCPENNAISETDFRNFQMVPLDPKGTQIDPNWIR 982
 Db 495 adrdgvgdvcdqddfadkvdkidvcpenaevtltdfratqvtvldpegdaqldpnwvl 554
 QY 983 HOGKELVQANSDDPGIAGVDFEGSVDFTFYVNTDRDDYAGVFGVQSSSRFYVVMW 1042
 Db 555 nggrelvtmnsdpglavgytafngvdfegtfnvntvtdddyagfigyqsdssfyvmmw 614
 QY 1043 KQVTQTYWEDQPTRAYGYGSVSLKVVNSTGTGHEHLRNALWHTGTPGQVRTLWHDPRNI 1102
 Db 615 kqmeqtywqanpfravaepgiqkavkstgpggeqlrnalwhtgdesqrvllwkdpnrv 674
 QY 1103 GWKDYTYRWHLTTHRPKTYIRVLVHEGKQVMADSGPIYDQTYAGGRLGLVFESQEMVYF 1162
 Db 675 gwkdksyrfwqlhrpqvgvyrfrfyegpelvadnsnvltdtmrggrlgvfcfsqenlww 734
 QY 1163 SULKYECRD 1171
 Db 735 anlryrcnd 743

RESULT 10
 AAM93335
 ID AAM93335 standard; Protein; 762 AA.
 XX

AA093335;
06-NOV-2001 (first entry)
Human polypeptide, SEQ ID NO: 2869.
Human; full length cDNA; cDNA synthesis; oligo-capping.
Homo sapiens.
EP1130094-A2.
05-SEP-2001.
07-JUL-2000; 2000EP-0114089.
08-JUL-1999; 99JP-0194496.
11-JAN-2000; 2000JP-0118774.
02-MAY-2000; 2000JP-0183765.
(HELI-) HELIX RES INST.
Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y,
Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H.
WPI; 2001-524255/58.
N-PSDB; AAK94255.
830 Primers useful for synthesizing full length cDNA clones and their
use in genetic manipulation -
Claim 8; SEQ ID NO 2869; 1380pp + sequence listing; English.
The invention relates to primers for synthesizing full length cDNA
clones. 830 cDNA molecules encoding a human protein have been
isolated and nucleotide sequences of 5' and 3' ends of the cDNA
molecules have been determined. Primers for synthesizing the full length
cDNA are useful for clarifying the function of the protein encoded by
the cDNA. The full length clones were obtained by construction of full
length enriched cDNA libraries that were synthesised by the oligo-capping
method. The primers enable the production of the full length cDNA easily
without any special methods. The present sequence is a polypeptide
encoded by a full length human cDNA of the invention.
Note: The sequence data for this patent did not form part of the printed
specification, but was obtained in CD-ROM format directly from EPO.
Sequence 762 AA;
Query Match 29.58; Score 1946.5; DB 22; Length 762;
Best Local Similarity 52.28; Pred. No. 4.1e-112;
Matches 360; Conservative 78; Mismatches 183; Indels 69; Gaps 11;
QY 537 VQEROMC---NKRSCPVDGCLSNPCFCAQCS---SPFDGSMSCGSCPVGFLNGTHCED 590
DB 69 lmeqcvcfhegrs----hcsnpncfrgvdcmeyeyp--gyrcgpcpglqngthcsd 122
QY 591 LDECALVPDICTSKVPRCVNTOPGFHCLPCPPRYRGNQPVGVGLEAAKTEKOVCE--- 647
DB 123 lnecaha-dpcfpss---clntmpgfceacprgkgtgvgidyaraskqvcndid 178
QY 648 -----PENPCDKDTHN-CHKHAEC 665
DB 179 ecndngnggcdpnsictntvsgfkcgpcrlgflngsgcglpartchspahspchihaic 238
QY 666 IYLGHFSPMYKCEQTYAGDGLICGEDSLDGLWPNLNLVCATNATVHC1KDCNCPHLPN 725
DB 239 lf---erngavscnvgagngnvcgtcdtdidydpdqlpcmdn-nkhckqdcclltpn 294
QY 726 SQGEDFDKDGIGDADDDDDNDGVTDERDNQQLFNPRQADYDKDEVDGDCNCPYVHNP 785
DB 295 sggedadndvgdcdaddadgiknvedncrlfpnkqgqnsdtdsfgdacdcnncpvpnn 354

QY 786 AQIDTNNEGDACSVDIDGDDVFNERDNCPPVYNTDQDTDGDGVDGDCNCPVLVHNP 845
DB 355 dqktdngsgedacndndvddgipnglancpkvpnpqlctdrdedvgdacsccpemsapt 414
QY 846 QTDVNDLVGDCDNNEDIDDDHQNNDNCPYISNAQADHRDQGGQDADPDNDNGV 905
DB 415 qtdadsdlvgdcdtnedsdgdgtdktdncpqlpnssqldsdndglgdecddndgi 474
QY 906 PD---DRDNCRLVFNPDQEDLDGDRGDIKDDFDNDNIPDIDDVCPENNAISTDFRN 961
DB 475 pdyppdpdncrlvpnpnqkdsdngvgvceddldndavvdpdvdpesaevtldfra 534
QY 962 FQVPLDPKGTQIDPNMVIIRHOGKELVQTANSPDGIAGVDFEGSVDFSGTFYVNTDRD 1021
DB 535 yqtvvldpegdaqidpnwvlnqgmewtmsdpglavgytafngvdfegtfhvtvtd 594
QY 1022 DDYAGVFVGYSRSSRFYVMKQVTOYWEQDTPRAYCYSGVSLKVVNSTTGTGHEHNA 1081
DB 595 ddyagflfsyqdsgrfyvwmkqteqtwqatpfravaqpglqkavtsvsgpgehlrna 654
QY 1082 LWHGTNTPGOVRTLWHDPRNIGWKDYTAIRWHLTHRPKTYIRVLVHKGQVMADSGPIY 1141
DB 655 lwhghtpdpqvrllwtddprnvgwrktsyrgwllhrpqvgvyrkvlyegpqlvadsgvli 714
QY 1142 DQYTAGGRLGFLVFSQEMVYFSDLYKECRD 1171
DB 715 dtsmrgrglgvfcfsqenilwslqyrnd 744
RESULT 11
AAR56248
ID AAR56248 standard; Protein; 889 AA.
XX AC AAR56248;
XX DT 20-JAN-1995 (first entry)
XX DE Xenopus thrombospondin-4.
XX KW Thrombospondin-4; transgenic animal.
XX OS Xenopus laevis.
XX PN W09413794-A.
XX PD 23-JUN-1994.
XX PF 03-DEC-1993; 93WO-US11725.
XX PR 04-DEC-1992; 92US-0985296.
XX PA (BGHM) BRIGHAM & WOMENS HOSPITAL.
XX PI Lawler JW;
XX DR WPI; 1994-217874/26.
XX DR N-PSDB; AAK66454.
XX PT Human thrombospondin-4 - made by recombinant DNA methods, useful
in experimental and therapeutic protocols
XX PS Disclosure; Page 45-47; 66pp; English.
XX CC cDNA encoding X. laevis thrombospondin-4 was cloned by PCR using
degenerate primers (AAQ66456-57) based on conserved regions of various
thrombospondins. The cDNA sequence (AAQ66454) and corresponding amino
acid sequence (AAR56248) were determined.
XX SQ Sequence 889 AA;
Query Match 28.88; Score 1899; DB 15; Length 889;
Best Local Similarity 51.98; Pred. No. 4.2e-109;

Matches 348; Conservative 79; Mismatches 184; Indels 60; Gaps 7;	
QY 549	PVDCCLSNPCFPGAQCSFPDGSWSCGSCPVGLGNGTHCEDLDECALVPDIFSTSKVP 608
Db 215	pkprcdatscfrgvrclid-teggfqcpcpegtygtgngvictdvdecrlnp--cflg---v 268
QY 609	RCVNTQPFHCLPCPPRYRGQNPVGVGLEAAKTEKQV----- 645
Db 269	rcintspgfkcespppytstlqiginfakqkvctdtneengrngcstnsalcin 328
QY 646	-----CEPENPKDKTHNCHKHAEICYLGHFSDFMYKCECQTGY 684
Db 329	tmgscfrcgckpgyvgvqikgckpekscrbhgqnpchasaqc--seekdgvctctsvgw 385
QY 685	AGDGLICGESDLDGWNILNLCATNATYHCIDKNCPLHNSQGEDFDKDGIGDADDDDD 744
Db 386	agngylcgktdldgydpdealpcpdk---nckkdcnvypnsqgedtknigdadeda 442
QY 745	DNDGVTDEKNCOLLFPNPRADYDKDEVGRDCNCPVYVHNPAAQIDTDNNGEGDACSVDID 804
Db 443	dggilneqncvlaanaidqnsdqdlifgdacdnclrlhndqrtdndgkgdaccddmd 502
QY 805	GDDVFNERNDCPVYVYNTDQRTDGDGVGDHDCNCPVHNPDPDQTDVNDLVGDQCDNNEI 864
Db 503	gdgiknildncqrvpndvqkdkdgdvgdldcscpdilnpnqgsdiddndlvgsdcdtnqds 562
QY 865	DDGQHQNQNCNPYISNANQADHGRDGGACDPPDDNDGVPD-----DRDNCRLVFNPDQ 920
Db 563	dgdghqdstncptvinsnqltdkdgigdecddddndgipdvppgpnccklvnpnpg 622
QY 921	EDLDGGRGICDKDDFNDNIPDIDVCPENNAISFDRFNFMVPLDPKGTQIDPNWV 980
Db 623	edndndvgvceadfqdvtidrvcpenaeltldfrayctvvlidpegdqiopnw1 682
QY 981	IRHOGKELVQTANSDPGIAVGDFEFGSGTYVNTDRDDYAGVFGYQSSRFYV 1040
Db 683	vlnqgmseivqtmsdpglavgytafngvdfegtfhvntmdddyagfifgqgsssfyv 742
QY 1041	MWQVQTYWEDQPTRAYGYSGLKVVNSTTGTGHEHLRNALWHTGNTPGQVTLWHDPR 1100
Db 743	mwaqtedyqatpfraavepglqkavksksgpgehlrnalwhtgtdndqvrllwkdp 802
QY 1101	NIGWKDYATYRWHLTHRPKTYIRVLVHEGKQVMADSGTYDQYAGGRLGLFVFSQEMV 1160
Db 803	nvgwkdkvsyrfqlhrpqvgylrarfyegtelvadsqvtvdtmrggrlgvfcfsqeni 862
QY 1161	YFSDLKYECRD 1171
Db 863	lwnlkyrcnd 873
RESULT 12	
AAR56249	
ID	AAR56249 standard; Protein; 961 AA.
XX	
AC	AAR56249;
XX	
DT	20-JAN-1995 (first entry)
XX	
DE	Human thrombospondin-4.
XX	
KW	Thrombospondin-4; transgenic animal.
XX	
OS	Homo sapiens.
XX	
PN	WO9413794-A.
XX	
PD	23-JUN-1994.
XX	
PF	03-DEC-1993; 93WO-US11725.
XX	
PR	04-DEC-1992; 92US-0985296.
XX	

PA	(BGHM) BRIGHAM & WOMENS HOSPITAL.
XX	
PI	Lawler JW;
XX	
DR	WPI; 1994-2117874/26.
DR	N-PSDB; AAQ66455.
XX	
PT	Human thrombospondin-4 - made by recombinant DNA methods, useful in experimental and therapeutic protocols
XX	
PS	Disclosure; Page 49-51; 66pp; English.
XX	
CC	cDNA encoding X. laevis thrombospondin-4 was cloned by PCR using degenerate primers. The predicted amino acid sequence was searched for regions of low codon degeneracy. PCR primers 89PCR (AAQ66458) and 90PCR (AAQ66459) were based on 2 such regions, and were used to isolate the human homolog of thrombospondin-4 from a heart cDNA library. Encoding sequences can be used to produce recombinant thrombospondin-4, or to breed transgenic animals carrying or lacking the thrombospondin-4 gene.
XX	
SQ	Sequence 961 AA;
Query Match 28.6%; Score 1890; DB 15; Length 961;	
Best Local Similarity 53.1%; Pred. NO. 1.7e-108;	
Matches 355; Conservative 71; Mismatches 185; Indels 58; Gaps 9;	
QY 549	PVDCCLSNPCFPGAQCSFPDGSWSCGSCPVGLGNGTHCEDLDECALVPDIFSTSKVP 608
Db 286	pprdsanpcfrgvcctdsrdg-fqgpcpegtygtgngitcidvdeckyhp--cypg---v 339
QY 609	RCVNTQPFHCLPCPPRYRGQNPVGVGLEAAKTEKQVC-----EPEN----- 650
Db 340	hcinslpgfrdcacpvgftgpmvgvgvgsfaksnkvctdidecngacvpslcvntlg 399
QY 651	-----PCK-----DKTHNCHKHAEICYLGHFSDFMYKCECQTGYAG 686
Db 400	syrcgpcpgytdqirgckvncrnrcnlpncsvnaqcl---eergdvtcvcgvwgag 456
QY 687	DGLICGEDSLDGMPLNLVLCATNATYHCIDKNCPLHNSQGEDFDKDGIGDADDDDD 746
Db 457	dgyicgkdvdisypdeelpcsar---nckkdcnvypnsqgedadrdgigdcadadag 513
QY 747	DGVTDEKNCOLLFPNPRADYDKDEVGRDCNCPVYVHNPAAQIDTDNNGEGDACSVDID 806
Db 514	dgilneqncvlihnvdqnsdkdfigdacdnclslvnnndqktdgdgrgdacddmdgd 573
QY 807	DVFNERNDCPVYVYNTDQRTDGDGVGDHDCNCPVHNPDPDQTDVNDLVGDQCDNNEI 866
Db 574	giknildncpckfnpdrkdgdvgvadscdcpdvsnpnqgsdvndlvgsdcdtnqdsdg 633
QY 867	DGHQNQNCNPYISNANQADHGRDGGACDPPDDNDGVPD-----DRDNCRLVFNPDQED 922
Db 634	dghqdstncptvinsnqltdkdgigdecddddndgipdlvppgpncrlvnpnqad 693
QY 923	LDGDRGDIKDDFNDNIPDIDVCPENNAISFDRFNFMVPLDPKGTQIDPNWVIR 982
Db 694	snsdvgdicesdfddqdvldrvdcpenaevctifrayctvvlidpegdaqldpnwv1 753
QY 983	HOGKELVQTANSDPGIAVGDFEFGSGTYVNTDRDDYAGVFGYQSSRFYVVMW 1042
Db 754	nqgmseivqtmsdpglavgytafngvdfegtfhvntqtdddyagfifgqgsssfyvwm 813
QY 1043	KQVQTYWEDQPTRAYGYSGLKVVNSTTGTGHEHLRNALWHTGNTPGQVTLWHDPRNI 1102
Db 814	kqteqywdatpfraavepglqkavksktpgehlrnalwhtgtdsdqvrllwkdsrnv 873
QY 1103	GWKDYATYRWHLTHRPKTYIRVLVHEGKQVMADSGPIYDQTVAGGRLGLFVFSQEMVYF 1162
Db 874	gwkdkvsyrfqlhrpqvgylrarfyegselvadsqvtidtmrggrlgvfcfsqeni1w 933
QY 1163	SDLKYECRD 1171


```

Db 934 snlkyrcnd 942
RESULT 13
ABB65649
ID ABB65649 standard; Protein; 1024 AA.
XX AC ABB65649;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster polypeptide SEQ ID NO 23739.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX DR WPI; 2001-656860/75.
XX DR N-PSDB; ABL09752.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PS interactions -
XX PS Disclosure; SEQ ID NO 23739; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins
XX CC (ABB57737-ABB72072).
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 1024 AA;

Query Match 27.38; Score 1805; DB 22; Length 1024;
Best Local Similarity 34.08; Pred. No. 3.2e-103;
Matches 407; Conservative 134; Mismatches 354; Indels 302; Gaps 31;

Qy 77 IMRQKEGFLTAQKQDGKSGTLLALEGPGLSQ-ROFEIVSNGPADTLDTYIDWTRH 135
Dy 6 mlkhmsfll-----drkqrvtlidisangatesrnfelplnetati----- 48
Qy 136 VVSLVDGLADSQMKNVTVVAGETYSLVHVCGLDLSFDALDEPFYHLOAKFSRMVAKG 195
Dy -----rsiaigkskrithlvdcaksthhdid-----mmlakiyqmd 86
Qy 196 SARESHFRGLLQNVHLFENSVEDILSKGCGQG---QGAEINAISENTEFLRLGPHV-- 250
Dy 87 dpviklfrerkyphl--fdgmehslqancqkgnhrgmrlnrkitergeifprilg 144
Qy 251 -----TTEV-----GPSSRRPEV---CER----- 268

```

```

Db 145 ylpqymiyisawypifaeaknkrdvrgwvptiaregvvdhrhgevptdvgdipvl 204
Qy 269 --SCEELGNMVOELSLGLVNLVNPQSENLKRVSNQDNQFELWELIGPPKTRNMSACWQDGRF 326
Db 205 ngdceda--larslsdlalvklirevahrqaeiyl-----rm 242
Qy 327 FAENETWVVDSCITCT-CKFKFTICHQITC-PPATCASPSEVEGECC-----PSCLHSVD 379
Db 243 llen-----cagcknplttndnglrliepcrsanpcygvvecidsaagprcgh--- 289
Qy 380 GEEGWSFWAEWTQCSVTGCGSTQQRGRSCDVTNTCLGPSIQTRACLSLSKDKTRIRODGG 439
Db 290 -----cpigfldgksck-----pgv---tcahnmcypvgv--- 317
Qy 440 WSHNSPWSGCVTGVGNITRILCNRPVPMQMGKNCKGSGRETAKCQAGACPIDGRWSP 499
Db 318 -----chdtv-----ngaqc----- 328
Qy 500 WSPWSACTVTCAGGIRTRVCNSPEPOYGGKACVGDVQERQMCNKRSCPVGDGCLSNPCF 559
Db 329 -----scpagyegdgtrtcslnp-----cldtcp 353
Qy 560 PGAQC--SSFPDGSWSGSCSPVGLNGTHCEDLDECALVPDICFSTSKVPRCVNTOPGF 617
Db 354 sgaclqvgyp-yfhciscpmghevngtsrdmnecllydpc---delatcnlspgf 408
Qy 618 HCLPCPPRYRGNPVGVGLE--AAKTEKQVCEPENPCDKTHNCKHAECI----- 666
Db 409 qscpcvfgdthahgyfadyysvyrqclvdcertgfrfrcpehstcineigsrycq 468
Qy 667 -YLGHFSDPMY-----KCEQTGYAGDGLIGED 694
Db 469 chegyvtngtyclsdrrsvfmcpgdgtvcdnavclrmdnirhkhcnvgnwagnglicgrd 528
Qy 695 SLDLGNLNLVLCATNATYHCIDKNCPLHNSGOEDFKDGIGDACDDDDNDGVTDEK 754
Db 529 tdvsgwpdqaigc---pelrcqrncpklpnsqgddadldghngdgcddadgdnvnsqd 585
Qy 755 NCQLLFPNQADYDKBVGDRCDNCPYVHNPAQITDNNCEGDSACSVIDIGDDVFNERN 814
Db 586 ncwlaynteqlsdgdkvgdvcdncvlyknpqrqldtddedglgdecgdldndsipnaldn 645
Qy 815 CPYYVNTDQRTDGDGVGDHCDNCPLYVHNPDTVDNDLVGDQCDNNEIDDDGHQNNQD 874
Db 646 cpilpnpsqsdvndvgdgcncnlpnpdqkdrmdfvgdchrlddgdgdpvpsld 705
Qy 875 NCPYISNANOADHRRDGGDGDGDDDDNCGVPPDDRCRLVFNPDQEDLDGDRGDIKCD 934
Db 706 ncpmvsnsdqltdgdgtgdecddmddgdgipnykdnopkpnkqddfnrnkgdsced 765
Qy 935 DFNDNDIPDIDVCPENNAISETDFRNFMVPLDPKGTQTQIDPNWVIRHOGKELVQFANS 994
Db 766 dedvvgvngmdncpnsmihhtdfrtltqtlpdkglsgadpwnvvhangteivqtlns 825
Qy 995 DPGTAGVDFEFGSVDFSGTFYVNTDRDDDDYAGFFVFGQSSSRFYVVMKWQVQTYWEDQP 1054
Db 826 dpgiavgkdaifggvdfgtfyndtdtdddvagfvsygsykyvvyvkkgtqtywep 885
Qy 1055 TRAYGYSVSLKVVNNTGTGTEHLNALWHTGNTPGVVRTLWHPDRNWKDYATYRWHL 1114
Db 886 ftasaeppgiqlklnvstceggpmrnslwhgegnctdgarllwkdpknlawkertsyws 945
Qy 1115 THRPKTYIRVLVHEGQVMADSGPIYDQTYAGAGRLGLFVFSQBMYFSDLYKECRD 1171
Db 946 vhrpaiglrilqmhagnrlifdsngvfdstlkgrlgvfcfsqrmilwsnlqykenn 1002

```

```

RESULT 14
AAB58933
ID AAB58933 standard; Protein; 444 AA.
XX AC AAB58933;

```

XX 27-MAR-2001 (first entry)
DE Breast and ovarian cancer associated antigen protein sequence SEQ ID 641.
XX Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
KW neotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
KW antidiabetic; antinflammatory; antitumor; antiparasitic; anticonvulsant;
KW antibacterial; antifungal; antiparasitic; cardiant; immune disorder;
KW Addison's disease; allergy; autoimmune haemolytic anaemia;
KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
KW cardiovascular disorder; wound healing; neurological disease.
XX Homo sapiens.
XX WO200055173-A1.
XX 21-SEP-2000.
XX 08-MAR-2000; 2000WO-US05881.
XX 12-MAR-1999; 99US-0124270.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Ruben SM;
PI WPI; 2000-611515/38.
XX N-PSDB; AAF21836.
XX New human breast and ovarian cancer associated gene sequences and the
PT polypeptides encoded by these genes, useful in the prevention,
PT treatment and diagnosis of cancer, immune disorders, cardiovascular
PT disorders and neurological diseases -
XX Claim 11; Page 1080-1082; 1299pp; English.
XX Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
CC proteins AAB58711 - AAB59128. The DNA and protein sequences are
CC associated with breast and ovarian cancer. Included in the invention are
CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the
CC isolation and characterisation of the DNA and protein sequences of the
CC invention. The breast and ovarian cancer associated DNA, protein, agonist
CC or antagonist sequences exhibit cytostatic; immunosuppressive;
CC neotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
CC antidiabetic; antinflammatory; antitumor; antiparasitic and cardiant activity;
CC antibacterial; antifungal; antiparasitic and cardiant activity. The
CC polynucleotide and protein sequences are used in the diagnosis of cancer,
CC particularly breast and ovarian cancer. The nucleic acid sequences,
CC proteins, agonists and antagonists may also be used in the diagnosis,
CC prevention and treatment of immune disorders e.g. Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; cardiovascular disorders such as
CC myocardial ischaemia; wound healing; neurological diseases such as
CC cerebral anoxia and epilepsy; and infectious diseases.
XX Sequence 444 AA;
SQ

Query Match 21.6%; Score 1428; DB 21; Length 444;
Best Local Similarity 60.7%; Pred. No. 2 4e-80;
Matches 258; Conservative 47; Mismatches 116; Indels 4; Gaps 1;
QY 751 DEKDNCLLPNRAQYDKDEVRDNCNCPYVHNPAQIDFDNNGEGDACSVDIDGDVFN 810
Db 1 neqncvllhdvqrsdkdfigdacnclsvlxndgktdggrgdacddmdgdkn 60
QY 811 ERNCPYVYTDRTDGGVGHCDNCPLVHNPDQTDVNDNLVGCDDNEDIDDDGHQ 870
Db 61 ildncpfnpdrgdkdgdvgdacsdpvsnpgsdvndlvgsdcdnqsdgdghq 120

QY 871 NNQDNCPIYSNANQADHRDGGQDADCDPDDNDGVPD----DRNCLVFNPDQEDLDGD 926
Db 121 dtdncdptvinsagldtdkdgdgdecddddgipolvpvgpndcrlvnpagdsnsd 180
QY 927 GRDICKDDFDNDINIPDIDVCPENNAISSETDFRNFQVPLDPKGTQIDPNNVIRHQK 986
Db 181 gvgdicesfdqddqvldrdvcpenaevltldfrayqvtvldpegdaqdpnvvlnqgm 240
QY 987 ELVQTANSPIAGVDFEGSVDFSGTFYVNTDRDDYAGVFGYSSSRFYVMMKQVT 1046
Db 241 elvqtmnsdpglavgytafngvdfegtfhvntqtdddyagfifgydsssfyvmwkqte 300
QY 1047 QTYWEDQPTRAYGYSGLKLVNNTGTGTEHLRNLAWHTGTGQVRTLWHDPRNIGWKD 1106
Db 301 qtywqatpfraavaepgiklavsktgpgehlrnlswhgtdsdqvrllwkdrrnvgkd 360
QY 1107 YTAYRWHLTHRPKTYIRVLVHEGKQVMADSGPIYDQTYAGGRLGLFVSQEMVYFSDLK 1166
Db 361 kvsyrvflqhrpqvgvyrvfegselvadsgvtldttmrggrlgvfcfsqenilwslk 420
QY 1167 YECRD 1171
Db 421 yrcnd 425
RESULT 15
AAG73734
ID AAG73734 standard; Protein; 444 AA.
XX AC AAG73734;
XX DT 03-SEP-2001 (first entry)
DE Human colon cancer antigen protein SEQ ID NO:4498.
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; chromosome 5.
XX Homo sapiens.
XX WO200122920-A2.
XX 05-APR-2001.
XX 28-SEP-2000; 2000WO-US26524.
XX 29-SEP-1999; 99US-0157137.
XX 03-NOV-1999; 99US-0163280.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
PI WPI; 2001-235357/34.
XX N-PSDB; AAF33165.
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers -
XX Claim 11; Page 6296-6298; 9803pp; English.
XX AAF32943 to AAF37195 and AAG73514 to AAG77788 represent human colon
XX cancer-associated nucleic acid molecules (N) and proteins (P), where
XX the proteins are collectively known as colon cancer antigens. The colon
XX cancer antigens have cytostatic activity and can be used in gene
XX therapy and vaccine production. N and P may be used in the prevention,
XX diagnosis and treatment of diseases associated with inappropriate P
XX expression. For example, N and P may be used to treat disorders
XX associated with decreased expression by rectifying mutations or deletions
XX in a patient's genome that affect the activity of P by expressing
XX inactive proteins or to supplement the patients own production of P.
XX Additionally, N may be used to produce the colon cancer-associated Ps,
XX by inserting the nucleic acids into a host cell and culturing the cell

CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX
SQ Sequence 444 AA;

Query Match 21.6%; Score 1428; DB 22; Length 444;
Best Local Similarity .60.7%; Fred. NO. 2.4e-80;
Matches 258; Conservative 47; Mismatches 116; Indels 4; Gaps 1;
QY 751 DEKDNQCLLFNPRQADYDKDEYDCNCPYVHNPAQIDTNNEGDACSVDIDGDDVFN 810
Db :||||| : : : ||||| || ||||| : ||||| : ||||| : ||||| :
1 neqdcnclihdvqnrskdldfgdaccnclsvlxndqktdgdgrgdacddmdgdgikn 60
QY 811 ERDNCPPYVNTDQRTDGVGDHCDNCPLVHNPQTVDNDLVGDCDNNEDIDDDGHQ 870
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
61 lldncpkfnpnrdgkdggvgdcdscpdvsnpnqsdvndnlvgdscdtnqsdgghq 120
QY 871 NNQDNCPIYSNANQADHRDGGDADCPDDDDNDGVDPD----DRDNCRLVFPDQEDLDGD 926
Db :: ||||| : : : ||||| || ||||| : ||||| : ||||| : ||||| :
121 dstdncptvinsaqldtdkdglgdecddddndgipdvppgpcnclrvpnpagedsnsd 180
QY 927 GRGDICKDDFDNDTIPDIDVCPENNAISETDFRNFOMVPLDPKGTTOIDPNWVIRHOGK 986
Db ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
181 gvgdicesdfddqvdridrldvcpnaevlttdfrayqtvtvldpegdaqldpnwvvingm 240
QY 987 ELVQTANSOPGIAVGDFERGVSDFSGTFYVNTDRDDDYAGVFYGQSSSRFYVVMWQVT 1046
Db ||||| ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
241 eivqtmsdpglavgytafngydfegtfnvtqtdddyagfifgydsssfyvvmmwkate 300
QY 1047 QYWEQPTFRAYGSGVSLKVYNSTGTCEHLRNALWHTNTPGQVRTLWHDPRNIGWKD 1106
Db ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
301 qtywqatpfravaepgiglkavkstktpgehlrnsliwhtgdtsgdqvrllwkdsrnvgwk 360
QY 1107 YTAIRWHLTHRPKTYIRVLVHEGQVMADSGPIYDQTYAGGRLGLFVFSQEMVYFSDLK 1166
Db :||| | ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
361 kvsyrwflqhrpqvgyirvfyegselvadsgvtidttmrggrlgvfcfsqeniwsnlk 420
QY 1167 YECRD 1171
Db |||||
421 yrcnd 425

Search completed: August 9, 2002, 10:01:16
Job time: 262 sec


```

QY 61 DYIPPNADLSKIKIMROKGGFFLTAQLKODGKSRGTLTALALEGPGLSORQFEIVSNP 120
DB 61 DYIPPNADLSKIKIMROKGGFFLTAQLKODGKSRGTLTALALEGPGLSORQFEIVSNP 120
QY 121 ADTLDTLTYWIDGTRHVVSLDVLGADSQWKNVTVQVAGETYSLHVGCDDLIDSFALDEFFY 180
DB 121 ADTLDTLTYWIDGTRHVVSLDVLGADSQWKNVTVQVAGETYSLHVGCDDLIDSFALDEFFY 180
QY 181 EHLQAEKSRMYAKSARESHFRGLLQNVHLVFNSEVEDILSKKCGQOQGAENAISEN 240
DB 181 EHLQAEKSRMYAKSARESHFRGLLQNVHLVFNSEVEDILSKKCGQOQGAENAISEN 240
QY 241 TETLRGPHVTTEYVGPSSERPEVCERSCBELGNMVOELSLHLVNPQSENARVND 300
DB 241 TETLRGPHVTTEYVGPSSERPEVCERSCBELGNMVOELSLHLVNPQSENARVND 300
QY 301 NOFLMELIGGPPKTRNMSACWQDGRFFAENETWYVDSCTTCTCKFKTKICHQITCPPATC 360
DB 301 NOFLMELIGGPPKTRNMSACWQDGRFFAENETWYVDSCTTCTCKFKTKICHQITCPPATC 360
QY 361 ASPSFVEGECPCSLHSDVDEGEGSWPABWTCQSVTCGSGTOQRGRSCDVTNTCLGPSI 420
DB 361 ASPSFVEGECPCSLHSDVDEGEGSWPABWTCQSVTCGSGTOQRGRSCDVTNTCLGPSI 420
QY 421 QTRACSLSKCDTRIRQDGGWSHSPWSSCVTCGVGNITRIRLNCNPSVPQMGKCKGSG 480
DB 421 QTRACSLSKCDTRIRQDGGWSHSPWSSCVTCGVGNITRIRLNCNPSVPQMGKCKGSG 480
QY 481 RETKACQAPCIDGRWSPWSPACTVTCAGGIRERTRVCNPSPEPQYGGKACVGDVQER 540
DB 481 RETKACQAPCIDGRWSPWSPACTVTCAGGIRERTRVCNPSPEPQYGGKACVGDVQER 540
QY 541 QMCNRKSCPDGCLNCPGPAQCSSFPGDWSWCSGCPVGFNGTHCEDLDECALVPDI 600
DB 541 QMCNRKSCPDGCLNCPGPAQCSSFPGDWSWCSGCPVGFNGTHCEDLDECALVPDI 600
QY 601 CFSTSKVPRCVNTQGFHCLPCPPRYRGQNPQVGVGLEAKTEKQVCEPENCKDKTHNCH 660
DB 601 CFSTSKVPRCVNTQGFHCLPCPPRYRGQNPQVGVGLEAKTEKQVCEPENCKDKTHNCH 660
QY 661 KHAECIYLGHSFDPYKCBQCYAGDGLICGEDSDLDGWPNLNLVCATNATHYHCKNC 720
DB 661 KHAECIYLGHSFDPYKCBQCYAGDGLICGEDSDLDGWPNLNLVCATNATHYHCKNC 720
QY 721 PHLPNSGQEDFKDGLIGDADDDNDGVTDEKNCOLLFNPRQADYDKDEVDRCDNCP 780
DB 721 PHLPNSGQEDFKDGLIGDADDDNDGVTDEKNCOLLFNPRQADYDKDEVDRCDNCP 780
QY 781 YVHNPAQIDTDNNGEGDACSVDIDGDDVFNERNDCPYVYNTDQRTDGDGVGDHDCNCP 840
DB 781 YVHNPAQIDTDNNGEGDACSVDIDGDDVFNERNDCPYVYNTDQRTDGDGVGDHDCNCP 840
QY 841 VHNPDQTDVNDLVGDCQDNNEIDDDGHQNDQNCPIYISNANQADHDHDEGGDADCPDD 900
DB 841 VHNPDQTDVNDLVGDCQDNNEIDDDGHQNDQNCPIYISNANQADHDHDEGGDADCPDD 900
QY 901 DNDGVPDDRDNCRNFVNPDQEDLDGDRGDCDDFDNDNIPIIDDVCPENNAISETDPR 960
DB 901 DNDGVPDDRDNCRNFVNPDQEDLDGDRGDCDDFDNDNIPIIDDVCPENNAISETDPR 960
QY 961 NFOMVPLDPKGTQIDPNVNIHQKELVQTANSDFGIAVGFDEFGSDVFGTFVYNTDR 1020
DB 961 NFOMVPLDPKGTQIDPNVNIHQKELVQTANSDFGIAVGFDEFGSDVFGTFVYNTDR 1020
QY 1021 DDDYAGFVGYOSSRFYVMKQYOTYWEQDPTRAYGYSGLKLVNNTTGTGEHLRN 1080
DB 1021 DDDYAGFVGYOSSRFYVMKQYOTYWEQDPTRAYGYSGLKLVNNTTGTGEHLRN 1080
QY 1081 ALWHTGNTFCQVRLWHDPRNIGWKDYATYRWHLTHRPKTYGIRVLVHEGKQVMAADSGPI 1140
DB 1081 ALWHTGNTFCQVRLWHDPRNIGWKDYATYRWHLTHRPKTYGIRVLVHEGKQVMAADSGPI 1140

```

```

QY 1141 YDOTYAGGRGLGVFSQEMVYFSDLKYECRDI 1172
DB 1141 YDOTYAGGRGLGVFSQEMVYFSDLKYECRDI 1172
RESULT 2
US-08-313-288B-20
; Sequence 20, Application US/08313288B
; Patent No. 5750502
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M. and Avihu Klar
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/313.288B
; FILING DATE: January 5, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 40028-A-PCT-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0526
; TELEX:
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1170 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-313-288B-20

```

```

Query Match 64.0%; Score 4228; DB 1; Length 1170;
Best Local Similarity 61.6%; Pred. No. 0;
Matches 725; Conservative 170; Mismatches' 267; Indels 14; Gaps 6;
QY 1 MYWRL-VLLALWVPSTQAGHODKD-TTFDLFSINIRKKTIGAKQFQPGDPGVPAYRNV 58
DB 3 LAWGVFLMHVCTGNRIPESSGDNVDFIFELTGAARKSGRRLKVGPDSPSAFRIE 62
QY 59 RFDYPPNADLSKIKIMROKGGFFLTAQLKODGKSRGTLTALALEGPGLSORQFEIVSN 118
DB 63 DANLIPPVDDKFDQDLVAVRTEKGFLLASLURQMKTRGTLLALERKDHSGQVFSVSN 122
QY 119 GPADTLDLTYWIDGTRHVVSLDVLGADSQWKNVTVQVAGETYSLHVGCDDLIDSFALDEP 178
DB 123 GKAGTLDLTLVQGGKHVSVVEALLATGQWKSITLQVQEDRAQLYIDCEKMEAELOVP 182
QY 179 F---YEHQAESRMVYAKSARESHFRGLLQNVHLVFNSEVEDILSKKCGQOQGAEN 235
DB 183 IQSVTRDLASTARLAKAGVND-NFQGVLRNVFVGTTPEDILRNKGCSSSTVLLT 241
QY 236 AISENTEYLRGPHVTTEYVGPSSERPEVCERSCBELGNMVOELSLHLVNPQSENARVND 295
DB 242 L--DNNVYNGSSPAIRTNVIGKTKDLQAICGISCDELSSWVLELRGLRTIVITLQDSIR 299
QY 296 RVSNQDQFLWELIGGPPKTRNMSACWQDGRFFAENETWYVDSCTTCTCKFKTKICHQITC 355

```

```
Db 300 KYTEENKELANEARRP-----LCHNGVQYRNNEWTVDSCTECHCONSVTICKKVSVC 353
QY 356 PPATCASPVEGECPCSLHSDVEGEGSPWAEWTQCSVTGSGTQGRGRSDVTSNRC 415
Db 354 PIMPCSNATVPDCECCPRCWPSDSADGNSPWEWTSCTSCGNGIQGRGRSCLNNRC 413
QY 416 LGPSIOTRACLSKCDTRTODGGWHSWSPSSCSTCGVGNITRILCNLSPVPQMGKN 475
Db 414 EGSSVOTRCHIQEODKRFQDGGWHSWSPSSCSTCGVGNITRILCNLSPVPQMGKN 473
QY 476 CKGSGRETRACQACAPCIDGRWSPWSPWACTVTCAGGIRERPRVCSNPPQYGGKACVG 535
Db 474 CESEARETRACKDACPINGGWPSPWDICSVTCGGVQKRSRLCNNPTPQFGGDCVG 533
QY 536 DVQEROMCNKRSPVDCGLSNPCFPAGQSSFPDGSWSCGSPVGLNGTHCEDIDECA 595
Db 534 DVTENQICNKQDCPIDGCLSNPCFAGVKCTSYPDGSKWGCACPPGYSGNGIQCTDVECK 593
QY 596 LVPDICFSTSKVPCVNTQGFHCLPCPPRYRCNQPVGVGLEAAKTEKQVCEPENCKDK 655
Db 594 EVDPAFCNHNHGERCENTDGYNCLCPFRFTGSGQGVGQVGHATANKQVCKRPNCTDG 653
QY 656 THNCHKHAECIYLGHSFSDPMYKCECOTGYAGDGLICGEDSLDGLWPNLNLVCAFNATYHC 715
Db 654 THDCNNAKCNLYLGHVSDPMYKCECPGYAGNGIICGEDTDLGWNENLVCVANATYHG 713
QY 716 IKDNCPLHNSGDEQDKGIGDACDDDDNDGVTDEKNCQLLNPQADYDKDEVGDR 775
Db 714 KDCNCPNLPSNGEDYDKGIGDACDDDDNDKIPDRDNCPPHYNPAQYDYDRDDVGDR 773
QY 776 CNDPCVYVHNPAQIDTNNEGDACSVDIGDDVFNERNDCNCPYVNTDQRTDGDGVDHC 835
Db 774 CNDPCVYVHNPAQIDTNNEGDACAADIDGDLNERDNCQYVYVNDQRTDMDGVDQC 833
QY 836 DNCPLVHNPDQTDVNDLVDGDCDNNEDIDGHNQNDNCNPIYISANQADHRRDQGDGA 895
Db 834 DNCPLHNPDQDSDSDRIGDTCNNQDIDEDGHNQNDNCNPIYVNPANQADHRRDQGDGA 893
QY 896 CDPDNDGVDPPDRDNCRLVNPQEDLDGDRGDIKCKDDFDNDNIPDIDVCPENNATS 955
Db 894 CDHDDNDGTPDDKNCRLVNPQDQSDGDRGDAKDDFDHDSVPDIDVCPENNATS 953
QY 956 ETDFRFQVMPDLPKGTQTDPMWVIRHOGKELVOTANSDPGTAVGDEGSDGSGTFY 1015
Db 954 ETDFRFQVMPDLPKGTQNDPMWVIRHOGKELVQVNCNCPGLAVGDEFNVDGSGTF 1013
QY 1016 VNTDRDDYAGFVFGYQSSSRFYVVMKQVTOYFWDQTPRAYGSGVSLKVVNSTGTG 1075
Db 1014 INTERDDYAGFVFGYQSSSRFYVVMKQVTOYFWDNTPRAYGSGVSLKVVNSTGTG 1073
QY 1076 EHLRNALWHTGNTPGQVRLTWHDPNRIGWKDYATYRWHLTHRPKTYIRVLVHEGKOVMA 1135
Db 1074 EHLRNALWHTGNTPGQVRLTWHDPNRIGWKDYATYRWHLTHRPKTYIRVLVHEGKOVMA 1133
QY 1136 DSGPIYDQYAGRGLGLFVSQEMVYFSDLKYECD 1171
Db 1134 DSGPIYDQYAGRGLGLFVSQEMVYFSDLKYECD 1169
```

RESULT 3

```
PCT-US93-11725-2
; Sequence 2, Application PC/TUS9311725
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: HUMAN THROMBOSPONDIN-4
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield, & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: United States of America
```

```
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3 1/2 inch
; COMPUTER: IBM-compatible
; OPERATING SYSTEM: MS-DOS Version 3.3
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/11725
; FILING DATE: filed herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/985,296
; FILING DATE: 04-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: GATES, Edward R.
; REGISTRATION NUMBER: 31,616
; REFERENCE/DOCKET NUMBER: B0801/7005WO
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 889 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: yes
; PCT-US93-11725-2
```

Query Match 28.8%; Score 1899; DB 5; Length 889;

Best Local Similarity 51.9%; Pred. No. 4.5e-139;

Matches 348; Conservative 79; Mismatches 184; Indels 60; Gaps 7;

```
QY 549 PVDGCLSNPCFPAGQSSFPDGSWSCGSPVGLNGTHCEDLDECALVDPICFSTSKVP 608
Db 215 PRPCDATSCFRVRCID-TEGFGQCPPEGYTGNGVICTDYDECLNP--CFLG---V 268
QY 609 RCVNTQGFHCLPCPPRYRCNQPVGVGLEAAKTEKQV----- 645
Db 269 RCINTSPGFKCESPPGYTGIQIGINFQKQKQVCTDNECENGRNGGCTSNSLCIN 328
QY 646 -----CEPENCKDKTHNCHKHAECIYLGHSFSDPMYKCECOTGY 684
Db 329 TMSGFRGCKPGYVGDQIKGCKPEKSRHGNQNPCHASAGC-----SEBKGDVTCVSGW 385
QY 685 AGDLGICGEDSLDGLWPNLNLVCAFNATYHCIDNCPHLPNSQEDFDKDIGDADDD 744
Db 386 AGNGYLCGRDITDIDGYPDEALPCPKD---NCKDNCVYVNPNSQEDTDKDNIGDCEDA 442
QY 745 DNDGVYDEKDNCLLPNROADYDKDEVGRDNCNPIYVHNPAQIDTNNEGEGSACSV 804
Db 443 DGDGILNEQDNCVLAANIDQKNSDQIDFGDACDNCRLTLNNDQRTDNDGKGACDDMD 502
QY 805 GDDVFNERNDCNPIYVNTDQRTDGDGVDHCNCPVHNPDQTDVNDLVDGDCDNNEDI 864
Db 503 GDGKILNDCQRPVNDQKDGVDGDIICDSCPDIIINPQSDIDNLDVGDSCDNTQDS 562
QY 865 DDDGHNNDNCNPIYISANQADHRRDQGDGACPDNDNDGVPD---DRNCRCLVFNPDQ 920
Db 563 DGDGHQDSTDCNCPVINSQDLTDGIDGECDDDDNDGIPDTPVPPNCKLVNPGQ 622
QY 921 EDLDGDRGDIKCKDDFDNDNIPDIDVCPENNAISETDFERNFQVPLDPKGTQIDPNWV 980
Db 623 EDDNDNGVDGVCADFDQDTVIDRVDVCPENAEITLTDPRAYQTVVLDPGDAQIDPNWI 682
QY 981 IRHQKELVOTANSDPGIAVGDFEVSDFSGTFYVNTDRDDYAGFVFGYQSSSRFYV 1040
Db 683 VLNQGMETVQTMNSDPGLAVGYTAFNGVDFEGTFHVTMTDITDQYAGFYQSSSRFYV 742
QY 1041 MKQVTOYFWDQTPRAYGSGVSLKVVNSTGTGHLRNALWHTGTGQVRLTHDPR 1100
Db 743 MKQTEQTIWQATPFRVAVAEPPGLKAVKSKSGPGEHLRNALWHTGTDNQVRLWKDPR 802
QY 1101 NIGWKDYATYRWHLTHRPKTYIRVLVHEGKOVMAADSGPTIYDQYAGRGLGVFSQEMV 1160
Db 803 NVGKMKVSRWFLQHRPQVGYIRARFYETGELVADSGVTVDTTMRGRLGVFCFSQENI 862
```


Db 1 MTEENKELANELRRPP-----LCYHNGVOYRNNEETVDSCTECHCONSVTICKKVSCLP 54
QY 357 PATCASPVEGECPCSLHSDVEEGWSPWAEWTQCSVTGSGTQQRGRSCDVTNTCL 416
Db 55 IMPCSNATVPDECCPCRPWPSDADGGHSPWSEWTSCTSCGNGIQQRGRSCDLSNRCE 114
QY 417 GPSIOTRACSLKCDTRIQDGGWHSWSPWSSCVTCGVGNITRIRLNSVPVQMGKNC 476
Db 115 GSSVOTRCHIQECDKRFKQDGGWHSWSPWSSCVTCGDGVITRITNLCSPOMNQKPC 174
QY 477 KSGRETRACOGACPIDGRSPWSPWACTVTCAGGIRERTV 520
Db 175 EGREAETACKRADCPINGWGPWSPWDICSVTCGGGVOKRSRL 218

RESULT 8
US-08-313-288B-15
; Sequence 15, Application US/08313288B
; Patent No. 5750502
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M. and Avi Hu Klar
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/313.288B
; FILING DATE: January 5, 1995
; CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 40028-A-PCT-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0526
; TELEX:

INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 469 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-313-288B-15

Query Match 6.38; Score 415; DB 1; Length 469;
Best Local Similarity 25.18; Pred. No. 3.4e-24;
Matches 124; Conservative 32; Mismatches 150; Indels 188; Gaps 20;

QY 360 CASPSFV---EGECCPCLHSDVEEGWSPWAEWTQCSVTGSGTQQRGRSCDVTNTCL 415
Db 57 CLNTAFATQKSGGLQCPC-----RSPRWLSMTWAPCSVTCSEGSQRLRYRCVGNWQC 111
QY 416 LGP-----SIQTRACSLSKCDTRIQDGGWHSWSPWSSCVTCGVGNITRIRLNSVPV 469
Db 112 SGKVAPGTLEWLQACEQOCCPEM---GGWSGWGPWPCSVTCSKGTTRRRACNHAP 168
QY 470 QMGKNCKGSGRETRAC-----QGA----- 489

Db 169 KCGG-HCPGQAQESACDQOVCPTHGAWATGWPWTPCSASCHGPHPEKETRSRKCSAP 227
QY 490 -----PCPIDGRWSPWSPWACTVTCAGSIRERTVCSNP 524
Db 228 EPSOKPKPKPCGLAYEQRRTGLPPCPVAGWGPWGPVSPCPVTCGLGQTMERTCNHP 287
QY 525 EPQYGGKACVGDQERQCMNKR-SCPVDGCLS-----NPC-----FPGAQCSFDPDGSWSC 574
Db 288 VPQHGPPCAGDATRTHICNTAVPCVPDGDWDSWGEWSPCIRRNKMSISCQEIP-GQOSR 346
QY 575 GSCPVGFLNG-----THCEDLDECALVPD-----ICFSTSKYPRCVNTOPG 616
Db 347 GRTCRGPKFGRHRCAGQOQDRIHCYSIQHCLKGSWSWSTWGLC-----MPPCGPNPTR 401
QY 617 FH---CLPCPPRYRGNQPVGVGLEAAKTEKQVCEPENPCDKDTHNCHKHAECIYLGHFSD 673
Db 402 ARQLCTPLLPKY---PPTVSMVGEQGEKNVT-----FWGR 434
QY 674 PMYKCECOTGYAGDGLICGEDSDLDGWPNLNLVCAATNATYHCIKDNCPLHPSNGQEDFDK 733
Db 435 PLPRCE---ELQGGKLVVEE-----KRPLHVP----- 459
QY 734 DGIGDADDDDDND 747
Db 460 -----ACKDPBEE 468

RESULT 9
US-08-918-914-4
; Sequence 4, Application US/08918914
; Patent No. 5876963
; GENERAL INFORMATION:
; APPLICANT: Mitchell, Peter
; APPLICANT: Hutchinson, Nancy
; APPLICANT: Lawton, Michael
; APPLICANT: Magna, Holly
; APPLICANT: Yocum, Sue
; APPLICANT: Murry, Lynn E.
; TITLE OF INVENTION: HUMAN NUCLEOTIDE PYROPHOSPHORYLASE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304

COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/918.914
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:

ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0369
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:

INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 788 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

Qy 832 -----GDHCNCPVHNPDQTDVNDLVGDCDNE---DIDDDGHNN 872
Db 779 SVWTAGSPCTTC-----KCKNGRVCCSVDFECLQNN 810

RESULT 11

US-09-055-699-34

; Sequence 34, Application US/09055699

; Patent No. 6005088

; GENERAL INFORMATION:

; APPLICANT: Tsutomu, FUJIWARA

; APPLICANT: Takeshi, WATANABE

; APPLICANT: Masato, HORIE

; APPLICANT: Toyomasa, KATAGIRI

; TITLE OF INVENTION: HUMAN GENE

; NUMBER OF SEQUENCES: 42

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sughrue, Mion, Macpeak & Seas

; STREET: 2100 Pennsylvania Avenue, N.W.

; CITY: Washington

; STATE: D.C.

; COUNTRY: United States

; ZIP: 20037-3202

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/055,699

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/820,170

; FILING DATE:

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 293-7060

; TELEFAX: (202) 293-7860

; TELEX: 6491103

; INFORMATION FOR SEQ ID NO: 34:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 810 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-09-055-699-34

Query Match 5.9% Score 388.5; DB 3; Length 810;
Best Local Similarity 20.6% Pred. No. 8.3e-22;
Matches 208; Conservative 106; Mismatches 347; Indels 349; Gaps 50;

Qy 8 LALVW-----PSTQAGH-QDKDTFDLFSISNINRKTIGAKQFQDPDPGVPAYRFRFD 61

Db 5 LILVWFVCVCFARTVWGMDPDQLMDIVTELDLVNTLGLVAQVSGMHNASKAFLODIE 64

Qy 62 ----YIPVNDLSKITKIMRQEGFELTAQLKDGKSRGTLALLEGSLSQRFELVS 117

Db 65 REIHAPVHS-----EKLIQLFQNKSEFILTATVOQKSTSGVILSIR--ELEHSYFELES 118

Qy 118 NPGADTLDLYWIDGTRHVVLEVDGLADSQMKNVTVOAGETYSYLVHVGCDLIDSPALDE 177

Db 119 SGLRDEIRYHYIHNGKPRTEAL--PYRMADGQWHKVALSVASHELLHVDNCRIVERIDP 177

Qy 178 PFYELQAKSRMYVAGSARESHFRGLLQNVHLVPENSVEDILSKGCGOQGAENAI 237

Db 178 P--DTNLPPGPNLWLGQRNKHGLFKGIQDKIIF-----MPNGYITQCPNL 223

Qy 238 SENTETLRGLPHVTTEYVGPSSRRPEVCERSCELGNNVOELSGLHVL-----VNQP 290

Db 224 NHTCPT-----CSDFLSLVUGIMDLQELLAKWTAKLNYA 257

Qy 291 SENLKRYSDNDNOFLWELIGGPPKTRNKSACHQDQREFAENETWV-VDSCTTCTCKKFKTI 349
Db 258 ETRLSQLEN-----CHCEKTCQVSGLLRQDQSWVDGDCRNCCTCKSGAVE 303
Qy 350 CHQITCPTATCAS---PSFVEGECC---PSCLSHS---VDGE-----381
Db 304 CRRMSCPPLNCSPDSLPHVIAGQCKVCRPKCIYGGKVLABGQRLITKSCRECRGGVLVK 363
Qy 382 -----EGWSPWAE-----W-TQCSVTCGSG-T 401
Db 364 ITEMCPPLNCSEKDHILPENQCCRCVRGHNFCABGPGKCGENSECKNNWTKATCECKSGYI 423
Qy 402 QQRGRS-----CDVTSNTCLGPSIQTRACSLSKDT---RIRQDGGWHSWPSWSSCS 450
Db 424 SVQDSAYCEDIDECAAKMHYCHANTVCVNLPLGLYRCDQCPGYIRVDD-----FSC 475
Qy 451 V--TCGVGNTRIRLNCSPVPMGGKNCCKSGRETAKCQAGAPCPIDGRWSWPSWASCTV 508
Db 476 EHDECGSGQ-----HNCDENAICTNTVQGHSC-----502
Qy 509 TCAGGIRERTVCNS---PEPOYGGKACVGDVQERQMCNKRSCP-----VDGCL 554
Db 503 TCRPGYVNGTICRAFCEBEGCRYGG-TCVAP-----NKCVCPSGFTGSHCEKIDIBCS 554
Qy 555 SN--PCFPGAQCSFDPGSMSCGCPVGFGLNGTH-----CEDLDBCALVPDICTFSTSK 606
Db 555 EGIIECHNHSRCVNLPG--GWYHCECRSGFDDGTYSLSGESCIIDIDECALATHTCWNDS- 611
Qy 607 VPRCVNTQPFHCLPCP--PRYRGNQVGVGLE-----AAKTEK-QVCEPENCKKQKTH 657
Db 612 --ACINLAGGFDCIL-CPSPGSCSGDCEHGLKNGQVWTLKEDRCVSVC---CKDKGI 664
Qy 658 NCHKHAEICVLGHFSDPMYKCEQOTGYAGDLGICGSDLDGWFNLVLCATNATYHCIK 717
Db 665 FCRTA-----CDCNPSA-DLFCCPE-----CDTRVTSQCLD 696
Qy 718 DNCPLHPSNGQEDFDKDGIGDADDDDDNDGVTDKONCQLLFNPR-QADYDKDEVGDRC 776
Db 697 QNGHKLYRSG-----DNWTHSCQCCRLGEVD---CWPLTCPLNLSCEYTAILEGCC 746
Qy 777 DNCPIVHNPAQIDTDNNGEGDACSVDIDGDDVFNERNCPYVNTDQDRTDGDGV----- 831
Db 747 PRC-----VSDPLAD---NITYDIRKTC-----LDSYGVSRLSG 778
Qy 832 -----GDHCNCPVHNPDQTDVNDLVGDCDNE---DIDDDGHNN 872
Db 779 SVWTAGSPCTTC-----KCKNGRVCCSVDFECLQNN 810

RESULT 12

US-09-273-565-34

; Sequence 34, Application US/09273565A

; Patent No. 6166190

; GENERAL INFORMATION:

; APPLICANT: FUJIWARA, TSUTOMU

; APPLICANT: WATANABE, TAKESHI

; APPLICANT: HORIE, MASATO

; TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN

; FILE REFERENCE: Q-53599

; CURRENT APPLICATION NUMBER: US/09/273,565A

; CURRENT FILING DATE: 1999-03-22

; EARLIER APPLICATION NUMBER: 09/055,699

; EARLIER FILING DATE: 1998-04-07

; EARLIER APPLICATION NUMBER: 08/820,170

; EARLIER FILING DATE: 1997-03-19

; EARLIER APPLICATION NUMBER: JP 63410/1996

; EARLIER FILING DATE: 1996-03-19

; EARLIER APPLICATION NUMBER: JP 69163/1997

; EARLIER FILING DATE: 1997-03-05

; NUMBER OF SEQ ID NOS: 95

; SOFTWARE: Patentin Ver. 2.1

; SEQ ID NO 34

```

; LENGTH: 810
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-273-565-34

```

Query Match	5.9%	Score 388.5	DB 4	Length 810	
Best Local Similarity	20.6%	Pred. No. 8.3e-22			
Matches 208	Conservative	106	Mismatches 347	Indels 349	Gaps 50
QY	8	LALWVW-----PSTQAGH-QDKOTFDLFSISNIRKKTIGAKQFRGDPGPVAPVREVRFD	61		
DB	5	LILVWFVCVCTATVVGFGMDPDLQMDIVITELDIVNTLTGLVAQVSGMHNASKAFLQDIE	64		
QY	62	-----YIPPVNADLSKTIIMROKEGFLLTAQLQDKGSRGTLTALALEGPGLSQRQFEIVS	117		
DB	65	REIHAAPHVS-----EKLIQLFONKSEFTILATVQKPSKSTGVLSIR--ELEHSVFELES	118		
QY	118	NGPADYLLTYWIDGTRHVVSLDEVDGLADSGQKNVTVQVAGETVSLHVGCDLIDSFALDE	177		
DB	119	SGURDEIRYHYIHNGKRPTEAL-PYRMADQGMHVALVSASHULLHVDCCRNIYERVIDP	177		
QY	178	PFYEHQAQESRMVYKAVGSARESHFRGLLQNVHLVFNSEVEDILSKKCGOQOGAEINAI	237		
DB	178	P--DTNLPPOINLWLGORNKHGLFKGIIQDGKIIF-----MNGYITQCPNL	223		
QY	238	SENTETLRLGPHVTVTVGSSERRPVCERSCEELGNMVQELSGHLVL-----VNQP	290		
DB	224	NHRCPT-----CSDFLSLVQGMIDQELLAKMTAKLNYA	257		
QY	291	SENLRKVSNDQFLWELIGGPKTRNMSACWQDGRFFAENETW-VDSCTTCTCKKFKTI	349		
DB	258	ETRLSQLE-----CHCEKTKQVSGLLYRDQDSSVDGDRMCTCKSGAVE	303		
QY	350	CHOITCPPATCAS---PSFVEGECC---PSCLHS---VDGE-----	381		
DB	304	CRMMSCPPLNCPSDLPVHIAGOCKVCVKPCYIGGVKVLAEGRILTKSCRECRGGVLVK	363		
QY	382	-----EGWSPWAE-----W-TQCSVTCGSG-T	401		
DB	364	ITEMCPPLNCSEKDHILPENQCCRVCRGHFACBPGKCGENSECKNNMTKATCECKSGYI	423		
QY	402	QQRGRS-----CDVTSNTCLGPSITQTRACLSKCDT---RLRQDGGWSHSPWSSCS	450		
DB	424	SVQGDSCAYCEDIDCAKMHYCHANTVCVNLPLYRLCDCVPGYIRVDD-----PST	475		
QY	451	V--TCGVGNTRIRLNCSPVPMGKKNCKSGRRTKACQACPCPIDGRWSPWSPWACTV	508		
DB	476	EHDECSGQ-----HNCDENAICTNTVQGHSC-----	502		
QY	509	TCAGGIRTRRVCS---PEPOYGGKACVGDVQERQMCNKRSCP-----VDGCL	554		
DB	503	TKPGYVGNGTICRAFCEGRYGG-TCVAP-----NKCVCPSGFTGSHCEKIDECSS	554		
QY	555	SN--PCFPFGAQCSSFPGWSCSGPCPVGFLNGTH-----CEDLDECALVPDICFSTSK	606		
DB	555	EGIIIECHNHSCVNLN-PGWYHCECRSGFDDGTYSLSGESCIDIDECALRTHTCWNSD-	611		
QY	607	VPRCVNTQPGFCHLPCP--PRYRGNQPVGVGLE-----AAKTEK-QVCEPENPCKDKTH	657		
DB	612	--ACINLAGGFDCL-CPSPGSCGDCPHEGLKHNGQVWTLKEDRCSVCS---CKDGKI	664		
QY	658	NCHKHAEICVLGHFSDPMYKCEQTGYAGGLIGCESDLDGWNLVLCATNATYHCIK	717		
DB	665	FCRRTA-----CDQNPSPA-DLFCFPE-----CDTRVTSQCLD	696		
QY	718	DNCPHLNPSGQEDFDKDIGDADDDDDNDGVTDKDNCCOLLFNPR-QADYDKDEVGDR	776		
DB	697	QNGHKLYRSG-----DNWTHSCQOCCRLSEVD---CWPLTCCNLSCETAILGECC	746		
QY	777	DNCPYVHNPAQITDNNNEGDAQSVDIGDDVDFVNERNCNCPYVNTDQRTDGDGV----	831		
DB	747	PRC-----VSDPCLAD---NITYDIRKTC-----LDSYGSVRLSG	778		

```

Qy      832  -----GDHCDCNCPLYVNPQTQVTDNDLYGDCQCDNNE---DIDDDGHQNN 872
          |   |   |   |   |   |   |   |   |   |   |   |   |
Db      779  SVWTMAGSPCTTC-----RCKNGRVCVCSDFECLQNN 810

RESULT 13
US-09-565-538-34
; Sequence 34, Application US/095655538
; Patent No. 6333404
; GENERAL INFORMATION:
; APPLICANT: FUJIWARA, TSUTOMU
; APPLICANT: WATANABE, TAKESHI
; APPLICANT: HORIE, MASATO
; TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN
; TITLE OF INVENTION: SKELETAL MUSCLE-SPECIFIC UBILQUITIN-CONJUGATING ENZYME
; FILE REFERENCE: Q-53599
; CURRENT APPLICATION NUMBER: US/09/565,538
; CURRENT FILING DATE: 2000-05-05
; PRIOR APPLICATION NUMBER: 09/273,565
; PRIOR FILING DATE: 1999-03-22
; PRIOR APPLICATION NUMBER: 09/055,699
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: 08/820,170
; PRIOR FILING DATE: 1997-03-19
; PRIOR APPLICATION NUMBER: JP 63410/1996
; PRIOR FILING DATE: 1996-03-19
; PRIOR APPLICATION NUMBER: JP 69163/1997
; PRIOR FILING DATE: 1997-03-05
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 34
; LENGTH: 810
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-565-538-34

```

```

Query Match      5.9%; Score 388.5; DB 4; Length 810;
Best Local Similarity 20.6%; Pred. No. 8.3e-22;
Matches 208; Conservative 106; Mismatches 347; Indels 349; Gaps 50;

Qy      8 LALVWV-----PSQAGH-QDKDTFDLFSTSNINRKTIGAKQPRGPDGPVAPRVREFD 61
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      5 LILVWFVCVARTVVGFGMDPDQLQMDIVTELDLVNTTGLGVAQVGMHNASKARFLDODIE 64
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy     62 ----YIPVNADSLKTKTKMRQEGFFLAQLKQDKSGRTLLIALSEGPLSGRQRFIVS 117
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     65 REITHAAPHVS-----EKLIQIQLFONKSEFTILATVQOKPSTSGVILSIR--ELEHSYFELES 118
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy    118 NGPADTLDLYWIDGTRHVVSLSDVGLADSQWKNVTVOVAGETYSLHVGCDLIDSFALDE 177
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    119 SGLRDEIRYHYIHNGKPRTEAL-PYRMADGQWHKVALVSASLHLLHWDNCNRIERVIDP 177
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy    178 PFYEHLQAEKRMVYKAGSARSHFRGLLQNHVLVFNSEVEDILSKGCGQOGGAENAI 237
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    178 P--DTNLPPGILWLGORNOKHGLFKGIIDGKIIF-----MPNGYITQCPNL 223
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy    238 SENTETLRLGPHVTEYVYGSSRRPCEVSRCEELNMVQELSLHLV-----VNQP 290
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    224 NHTCPT-----CSDFLSLVQGINMDQELLAKMTAKUNYA 257
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy    291 SENLKRYSDNQFLWELIGPPKTRNMSACWQDGRFAENETWV--VDSCTTCTCKKEFTI 349
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    258 ETRLSQLEN-----CHCEKTCQVSGLLLYRDQDSWVDGDHCRNCTCKSGAVE 303
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy    350 CHQITCPPATCAS---PSFVEGEC-----PCLHS-----VDGE----- 381
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    304 CRRMSCPPLNCSPDSLVPHIAGQCKVCRPKCIYGGKVLAEQGRILTKSCRECRGGVLVK 363
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy    382 -----EGWSPWAB-----W-TOCSVTCGSG-T 401
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    364 ITEMCPPLNGSEKHILPENQCCRVCRHGFACBPKGKNSCKNNKNTATCECKSGYI 423
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

402 QORGRS-----CDVTSNTCLGPSIOTRACSLSKDCT---RIRODGGWHSWSPSSCS 450
424 SVQGSAYCEDIDECAKMHVCHANTVCVNLPLGLYRCDVPGYIRVDD-----FSCT 475
451 V--TCGVGNITRIRLCHNSPVQMGKNCXGSGRETKACQACAPCPIDGRWSPWSPSACTV 508
476 EHDECGSQ-----HNCDENAICTNTVQGHSC----- 502
509 TCAGGIRERTVCNS---PEPOYGGKACVGDVQERQMCNKRSCP-----VDGCL 554
503 TKPGYVGVNGTICRAFCEGCRYGG--TCVAP-----NKCVCPSGFTGSHCEKIDDECS 554
555 SN--PCFPAGACSSFPDGSNCGSCPGVFLNGTH-----CEDLDECALVPDICEFSTSK 606
555 EGIIECHNSRCVNLPG--GWYHCECRSGFHDGTYSLSGSCIDIDEALRTHTCWNS- 611
607 VPCVNTQPGHCLPCLP--PRYRGNQPVGVGLE-----AAKTEK-QVCPEPNCKDKTH 657
612 --ACINLAGGFDCL-CFSGPSCSGDCPHEGGLKHNGQVWTLKEDRCVCS-----CKDGI 664
658 NCHKHAECIYLGHSDPMYCECQTAGYAGDGLIGEDSDLDGWNLVNLCATNATYICIK 717
665 FCRETA-----CDQNPASA-DLFCCPE-----CDTRVTSQCLD 696
718 DNPCLPNSGOEDFDKDGIGDADDDDDNDGVTDEKDNCLLFNPR-QADYDKDEVDGRC 776
697 QNGHLYRSG-----DNWTHSCQOCRCLEGEVD---CHPLTCPNLSCEYTAILEGCC 746
777 DNPYVHNPAQIDTNNNEGSDVSDIGDDVDFNERDNCPPYVNTQDQTDGCV----- 831
747 PRC-----VSDPCLAD--NITYDIRKTC-----LDSYGVSRSLG 778
832 -----GDHCDNCLVHNPDQTDVDNDLVGDQDCNNE---DIDDDGHQNN 872
779 SVWTMAGSPCTC-----CKNGRVCCSVDFECLQNN 810

RESULT 14
US-08-820-170A-37
Sequence 37, Application US/08820170A
Patent No. 5831058
GENERAL INFORMATION:
APPLICANT: Tsutomu, FUJIWARA
APPLICANT: Takeshi, WATANABE
APPLICANT: Masato, HORIE
APPLICANT: Toyomasa, KATAGIRI
TITLE OF INVENTION: HUMAN GENE
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sughrue, Milon, Zinn, Macpeak & Seas
STREET: 2100 Pennsylvania Avenue, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: United States
ZIP: 20037-3202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/820,170A
FILING DATE:
CLASSIFICATION: 536
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 293-7060
TELEFAX: (202) 293-7860
TELEX: 6491103
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 816 amino acids
TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-820-170A-37

Query Match 5.5%; Score 361.5; DB 2; Length 816;
Best Local Similarity 20.1%; Pred. No. 1.1e-19;
Matches 204; Conservative 126; Mismatches 354; Indels 333; Gaps 49;

QY 2 VRLVLVLLALWVWSTQAGHQDKDTTDFLSISINRRTIGAKQRPDPGPVAYRFRVFD 61
DB 19 VW-----GLGVDPSSLQ-----IDVLELGESTGTGVPGLHNGTKAFLDQTP 64
QY 62 YIPVNNADLSKITKIMROKEGFLTAQLKODGKSRGTLALALEPGLSQROFELVSNQPA 121
DB 65 RSIAKASTATAEQFFQKLRNKHETILVTLKQTHLNSVILSIH--HLDHYLELESSGHR 122
QY 122 DTLDTLWIDGTRHVSLEVDGLADSQMKNVTVOVAGETYSLAVGCDLIDSFALDEPFYE 181
DB 123 NEVRLHYRSGSHRPHTTEVPFVILADDKWHKLSAISASHLILHIDCNKI-----YE 173
QY 182 HLQAEKS-----RMVYKGSARESHFRGLLQNVHLVFNSEVEDILSKGCGQCGAEI 234
DB 174 RVVEKPSDPLGTTFWLQORNNAHGYFKGIMQDVQLLV-----MPQGFIAQC 221
QY 235 NAISENTETLRLGPHVTTEYVGPSSERRRPEVCERSCELGNNVQBELSGLHLVNPQSEN 294
DB 222 PDLNRTCTP-----CNDPFLGLVQKIMELQDILAKTSAKL 255
QY 295 -----KRVSNQNLFWELIGPPKTRNMSACWQDGRFFAENETWVVDSCITCTCKKFKTIC 350
DB 256 SRAEQRMRLDQCYCE-----RTCTMKGTTRYEFESW-IDGCKNCTCLNGTIOC 303
QY 351 HQITCPCATC---ASPSFVEGECPCSLHSDGEGHSPWAEWTOCSVTCSGSGTOQRGRS 407
DB 304 ETLICPNPDCPLKSAIAYVDKCC-----KEKSIK---QFGRT 340
QY 408 -CDVTSNTCLGPSIOTRACSLSKDCTRIRODGGWHSWSPSSCSVTGCVGNITRIRLNS 466
DB 341 YFEGERTVYSSS---GVCVLYECKDQ-----TMKLV-- 369
QY 467 PVQMGKNCXGSGRETKACQACAPCPIDGRWSPWSPSACTVTCAGGIRERTVCNSPEP 526
DB 370 -----ESSGCPALDC-----PESHOITLSHSCC----- 392
QY 527 QYGGKACVGDVQBROMCNKRSCPVDGCLSNPCFPAGACSSFPDGSWSCGSCPVGFLG--- 583
DB 393 ---KVCXG---YDFCSE-----HNCMENSICRNLDRA-VC-SCROGFRLRE 433
QY 584 NGTHCEDLDECALVPDICEFSTSKVPRCVNTQPGFHLCPRPYRGNQPVGVGLEAAKTEK 643
DB 434 DNAYCEDIDECAEGRHYCRENT---MCVNTPGSFMCI-CKTGY-----IRIDD 477
QY 644 QVCEPENPCDKTHNCHKHAECLYL--GHFSDPMYKCECOTGVAGDGLIGEDSDLDGWP 701
DB 478 YSCTEHDECITNQHNCDENALCFNTVGGH-----NCVCRPGYTGNGTTC-KAFCKGCR 530
QY 702 NLNLVLCATNATYHCIKDNCPLHPSGOEDFDKDGIGDADDDDD--NDGVT--DEKDNK- 756
DB 531 NNGGACIAANV---CA---CP-----QGTGPGSCETDIDECSDGFVQCDSRANCI 573
QY 757 -----OLLFNPROADYDKDEVGRDNCNCPYVHNPAIDTNNNEG----- 796
DB 574 NLPQWYHCECRDGYHDNGMFSPS-----GESCE-----DIDECGTGRHS 613
QY 797 --DACSVDDIDGDDVFNERNDCNCPYVNTDQRTDGDGVGDH-----CDNC 838
DB 614 ANDTICFNLDDG--YDCR--CPHGKNC-TGDCIHDGKVKHNGQIWLNDRCVSCSQNG 668
QY 839 PLVHNPDQTDVNDLVG-----DQCD-----NNEDIDDDHQNNDNCNCPYISANQADHRRD 891
DB 669 FVMCRMVCDENPTVDLFCCPECDPRLSSQCLHQNETLYNSGDTWVQNCQC---RCL 725

QY 892 QGDA-----C-DPDDNDGVPDRDNCRLVFNPDQEDLDGDRGDIKDDFDNDNI 941
Db 726 QGEVDCWPLPCPDVECEFSILPENECCPRCVTDPCQADTI---RNDITKTCLDEMNV 779

RESULT 15

US-09-055-699-37
Sequence 37, Application US/09055699
Patent No. 6005088

GENERAL INFORMATION:

APPLICANT: Tsutomu, FUJIWARA
APPLICANT: Takeshi, WATANABE
APPLICANT: Masato, HORIE
APPLICANT: Toyomasa, KATAGIRI
TITLE OF INVENTION: HUMAN GENE
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
STREET: 2100 Pennsylvania Avenue, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: United States
ZIP: 20037-3202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/055,699
FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/820,170
FILING DATE:

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 293-7060
TELEFAX: (202) 293-7860
TELEX: 6491103

INFORMATION FOR SEQ ID NO: 37:

SEQUENCE CHARACTERISTICS:

LENGTH: 816 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-09-055-699-37

Query Match 5.5%; Score 361.5; DB 3; Length 816;
Best Local Similarity 20.1%; Pred. No. 1.1e-19;
Matches 204; Conservative 126; Mismatches 354; Indels 333; Gaps 49;

QY 2 VNRLLVLLVWVSTQAGHQDQDTTDLFSISNINKRTIKAKQFRGPDGFPAYREVRED 61
Db 191 VW-----GLGVDFSLQ-----IDVLTELELGESTTGVQVPCVCLHNGTKRAFLFQDTP 64
QY 62 YIPVNAVDDLSTKTKIMROKEGFFLTAQLKQDGKSGRTLLALEGPGLSORQFEIVSNGPA 121
Db 65 RSIKASTATAEQPFQKLRNKHFEITLVTLKQTHLNSGVLSIH--HLDHRYLELESSGHR 122
QY 122 DTLDLYTWIDGTRHVVSLVDVGLDSQWKNVTVOVAGETYSLVHVGCDLIDSFALDEPFYE 181
Db 123 NEVRLYRSGSHRPHTEVPFYLLADKWHKLSAISASHILHIDCNKI-----YE 173
QY 182 HLQAEKS-----RMYVAKGSARESHFRLLQNVHLVFNSEVEDILSKGCGQOGAEI 234
Db 174 RVVEKSTDLPLGTFTEWLGORNNAHGYFKGIMQDVQLLV-----MPQGFIAQC 221
QY 235 NAISENFTELRGLPHVTTEYVGPSSRRPVCERSCEELGNMVQELSGHLVNLQPSNL 294
Db 222 PDLNRTCTP-----CNDFHGLVQKIMELQDILAKTSAKL 255
QY 295 -----KRVSNNDNQFLWELIGPPKPNRNSACWQDGRFAENETWVVDSCITCTCKKFKTIC 350

Db 256 SRAEQRMNRLDQCYCB-----RCTMTKGTYYREFESW-IDGCKNCTCLNGTIQC 303
QY 351 HQITCPPATC---ASPSFVEGECPCSLHSVDGEGWSPMAEWTCQSVTCGSGTQGRGRS 407
Db 304 ETLICPNPDCLKSALEYVDGKC-----KECKSIC---QFGRT 340
QY 408 -CDVTSNTCLGPSIQTRACSLSKCDTRIQRDGGGSHWSPSSCVTCGVGNITRILCNS 466
Db 341 YFEGERNYVSSS---GVCVLYECKDQ-----TMKLV--- 369
QY 467 PVPQMGKCKGSGRETAKACQCAPCPIDGRWSPWSPWACTVTCAGGIRERTRVCNSPEP 526
Db 370 -----ESSGCPALDC-----PESHQITLSHSCC----- 392
QY 527 QYGGKACVGDVQERQMKRSCPVGDCLSNPCFPAGACSFDPGWSWSSCSCPVGFLG--- 583
Db 393 ----KVCKG---YDFCSE-----HNCMENSICRNLNDRA-VC-SCRDGFRALRE 433
QY 584 NGTHCEDLDECALVPDICEFSTSKVPRCVNTQGFHCLPCPPRYRGVQVGVGLEAAKTEK 643
Db 434 DNAYCEDIDECAEGRHYCRENT---MCVNTPGSFMCI-CKTGY-----IRIDD 477
QY 644 QVCEPENCKDKTHNCHKHAECIYL--GHFSDPMYKCEGQTVAGDGLICGEDSDLDGWP 701
Db 478 YSCTEHEDECITNOHNCDENALCENTVGGH-----NCVCKPGYTGNGTTC-KAFCKDGR 530
QY 702 NLNLVCAATNATYHCKNDNCPHLPNSQEDFDKDGIGDADDDDD---NDGVT---DEKDN 756
Db 531 NGGACIAANV---CA---CP-----QGFTGPSCETOIDECSDGFVQCDSRANCI 573
QY 757 -----QLLENPRQADYDKDEVDRCDNCPYVHNPAQIDTNNNEG--- 796
Db 574 NLPGWYHCECRDGYHDNGMFSPS-----GESCE-----DIDEGCTGRHSC 613
QY 797 --DACSVDIDGDDVFNERNDCPVVYNTDQDQTDGCGVGDH-----CDNC 838
Db 614 ANDTICFNLDDG--YDCR--CPHGKNC-TGDCIHDKVKHNGOIWYLENDRCSVCSCQNG 668
QY 839 PLVHNPDDQTDVNDLVG-----DQCD---NNEDIDDDGHQNNQDNCPYISNANQADHRDG 891
Db 669 FVMCRRMVDCBNPTVDLFCCPECDPRLSSQCLHQNGETLYNSGDTWVQNCQOC---RCL 725
QY 892 QGDA-----C-DPDDNDGVPDRDNCRLVFNPDQEDLDGDRGDIKDDFDNDNI 941
Db 726 QGEVDCWPLPCPDVECEFSILPENECCPRCVTDPCQADTI---RNDITKTCLDEMNV 779

Search completed: August 9, 2002, 10:01:36
Job time: 162 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2002, 09:59:24 ; Search time 29.54 Seconds
(without alignments)
3812.346 Million cell updates/sec

Title: US-09-822-682-2
Perfect score: 6605
Sequence: 1 MWRLVLLALWVPSAQGH.....FVFSQEMVYFSLKAYECRDI 1172

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR71: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6570	99.5	1172	1 TSHP2	thrombospondin 2 p
2	5920	89.6	1172	2 A42587	thrombospondin 2 p
3	5481	83.0	1178	1 A39804	thrombospondin pre
4	4228	64.0	1170	1 TSHP1	thrombospondin 1 p
5	4189.5	63.4	1170	2 A40558	thrombospondin 1 p
6	1968	29.8	755	2 A44315	cartilage oligomer
7	1946.5	29.5	956	2 A57121	thrombospondin 3 p
8	1936.5	29.3	956	1 A46016	thrombospondin 3 -
9	1899	28.8	955	2 A45441	thrombospondin 4 -
10	1890	28.6	961	1 TSHP4	thrombospondin 4 p
11	1480	22.4	439	2 B44315	cartilage oligomer
12	580	8.8	128	2 S57955	thrombospondin 2 -
13	439.5	6.7	437	2 S05478	properdin - mouse
14	429	6.5	1444	2 T18856	angiogenesis inhib
15	424	6.4	1584	2 T00026	brain-specific ang
16	418.5	6.3	810	2 T10756	Nel-homolog protei
17	418	6.3	984	2 T00326	hypothetical prote
18	418	6.3	1522	2 T00028	brain-specific ang
19	416	6.3	469	1 S29126	properdin precursor
20	415	6.3	1074	2 JCS928	semaphorin F precu
21	405	6.1	957	2 T15976	hypothetical prote
22	399	6.0	788	2 T25061	hypothetical prote
23	384.5	5.8	1572	2 T00027	brain-specific ang
24	381	5.8	229	2 S57957	thrombospondin 1 -
25	343.5	5.2	2531	2 A46019	Notch-1 protein -
26	340.5	5.2	2531	2 S18188	notch protein homo
27	339.5	5.1	835	2 JP0076	nel protein - chlc
28	329.5	5.0	807	2 A38152	F-spondin - rat
29	329.5	5.0	2524	2 A35844	Xotch protein - Af

30	322	4.9	2471	2	A49128	cell-fate determin
31	318.5	4.8	2531	2	T31070	notch homolog - se
32	317	4.8	1360	2	T33922	hypothetical prote
33	312.5	4.7	861	2	A48825	Notch homolog Motc
34	311.5	4.7	2352	2	T30201	Notch homolog prot
35	308	4.7	803	2	A47723	F-spondin precursor
36	307	4.6	2871	2	A55624	fibrillin-1 precu
37	305.5	4.6	2703	1	A24420	notch protein - fr
38	302	4.6	2555	2	A40043	notch protein homo
39	300.5	4.5	2437	2	S42612	transmembrane prot
40	300	4.5	3002	2	A47221	fibrillin 1 precu
41	297	4.5	2871	2	A55567	fibrillin I - bovi
42	296	4.5	2139	2	A35672	crumbs protein - f
43	295	4.5	712	2	A45638	immunodominant mic
44	292	4.4	1558	2	C89114	protein C37C3.6a l
45	292	4.4	2167	2	T34395	hypothetical prote

ALIGNMENTS

RESULT 1
TSHUP2
thrombospondin 2 precursor - human
C:Species: Homo sapiens (man)
C>Date: 19-May-1995 #sequence_revision 03-Aug-1995 #text_change 13-Aug-1999
C:Accession: A47379; A42173
R:LaBell, T.L.; Byers, P.H.
Genomics 17, 225-229, 1993
A:Title: Sequence and characterization of the complete human thrombospondin 2 cDNA:
A:Reference number: A47379; MUID:94010892
A:Accession: A47379
A:Molecule type: mRNA
A:Residues: 1-1172 <LAB>
A:Cross-references: GB:L12350; NID:9307505; PIDN:AAA03703.1; PID:9307506
R:LaBell, T.L.; Milewicz, D.J.; Distche, C.M.; Byers, P.H.
Genomics 12, 421-429, 1992
A:Title: Thrombospondin II: partial cDNA sequence, chromosome location, and expression
A:Reference number: A42173; MUID:92217961
A:Accession: A42173
A:Molecule type: mRNA
A:Residues: 560-1172 <LA2>
A:Cross-references: GB:M81339
A:Experimental source: fibroblast
A:Note: sequence extracted from NCBI backbone (NCBIN:95091, NCBIP:95096)
C:Genetics:
A:Gene: GDB:THBS2; TSP2
A:Cross-references: GDB:128789; OMIM:188061
A:Map position: 6q27-6q27
C:Complex: homotrimer, disulfide linked
C:Function:
A:Description: participates in cell migration and adhesion, and in platelet aggregat
C:Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology
C:Keywords: beta-hydroxyasparagine; calcium binding; cell adhesion; glycoprotein; ti
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-1172/Product: thrombospondin 2 #status predicted <MAT>
F:319-377/Domain: von Willebrand factor type C repeat homology <VWC>
F:380-431/Domain: thrombospondin type 1 repeat homology <THR1>
F:436-492/Domain: thrombospondin type 1 repeat homology <THR2>
F:493-549/Domain: thrombospondin type 1 repeat homology <THR3>
F:553-588/Domain: thrombospondin type 1 repeat homology <EGF1>
F:652-691/Domain: EGF homology <EGF>
F:928-930/Region: cell attachment (R-G-D) motif
F:151-316-330-457-584-710-1069/Binding site: carbohydrate (Asn) (covalent) #status F
F:167-226/Disulfide bonds: #status predicted
F:266,270/Disulfide bonds: erythro-beta-hydroxyasparagine (Asn) #status predicted
F:612/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted

Query Match 99.5%; Score 6570; DB 1; Length 1172;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1167; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
QY 1 MWRLVLLALWVWSTQAGHODKDTTDLFSISINIRKTIAGAKQRPDPGPVPAFYRVRF 60
Db 1 MWRLVLLALWVWSTQAGHODKDTTDLFSISINIRKTIAGAKQRPDPGPVPAFYRVRF 60
QY 61 DYIPVNAADDLSKITKIMROKEGFFLTALQKODGKSRGTLALLEGPGISQRFVSNRP 120
Db 61 DYIPVNAADDLSKITKIMROKEGFFLTALQKODGKSRGTLALLEGPGISQRFVSNRP 120
QY 121 ADTLDLTWIDGTRHVSLEVDGLADSOQKNTVQVAGETYSLVHVCGLIDSFALDEFFY 180
Db 121 ADTLDLTWIDGTRHVSLEVDGLADSOQKNTVQVAGETYSLVHVCGLIDSFALDEFFY 180
QY 181 EHLQAEKSRMYVAKGASARESHFRGLLQNVHLVFNSEVEDILSKKCGQCGQGAENAISEN 240
Db 181 EHLQAEKSRMYVAKGASARESHFRGLLQNVHLVFNSEVEDILSKKCGQCGQGAENAISEN 240
QY 241 TETLRLGHVHTVTEYVGPSSRRPEYCSCEBELGNMVOELSLHLVNLQPSNKRVSND 300
Db 241 TETLRLGHVHTVTEYVGPSSRRPEYCSCEBELGNMVOELSLHLVNLQPSNKRVSND 300
QY 301 NOFLWELIGGPPKTRNMSACHQDGRFFFAENETWVDSCTTCTCKFKTICHQITCPPATC 360
Db 301 NOFLWELIGGPPKTRNMSACHQDGRFFFAENETWVDSCTTCTCKFKTICHQITCPPATC 360
QY 361 ASPSFVEGECPCSLHSDVGEKSWPWAETQCSVTCSGSGTQQRGRSCDVTSTNTCLGPSI 420
Db 361 ASPSFVEGECPCSLHSDVGEKSWPWAETQCSVTCSGSGTQQRGRSCDVTSTNTCLGPSI 420
QY 421 QTRACSLSKCDTRIQDGGSHSWPSSCVTCGVGNITRILCNPSVPQMGKNCKSG 480
Db 421 QTRACSLSKCDTRIQDGGSHSWPSSCVTCGVGNITRILCNPSVPQMGKNCKSG 480
QY 481 RETKACQAPCIDGRMSPWSPWACTVTCAGGIRERTVNCSPPOYGGKACVGDVGR 540
Db 481 RETKACQAPCIDGRMSPWSPWACTVTCAGGIRERTVNCSPPOYGGKACVGDVGR 540
QY 541 QMCNKRSPVDCGLSNPCPQAQCSFPGDWSGSCSPVGLGNTHGEDLDECALVPDI 600
Db 541 QMCNKRSPVDCGLSNPCPQAQCSFPGDWSGSCSPVGLGNTHGEDLDECALVPDI 600
QY 601 CFSTSKVPCRCVNTQPGFHLCPPRYRGVQVGLAETKQVCEPENCKDKTHNCH 660
Db 601 CFSTSKVPCRCVNTQPGFHLCPPRYRGVQVGLAETKQVCEPENCKDKTHNCH 660
QY 661 KHAECIYLGHSFSDPMYKCECOTGYAGDGLICGEDSLDGLWPNLNLVCATNATYHCIDNC 720
Db 661 KHAECIYLGHSFSDPMYKCECOTGYAGDGLICGEDSLDGLWPNLNLVCATNATYHCIDNC 720
QY 721 PHLPNSGQEDFKDGIIGDADDDNDGVTDEKNCQLLFPNQADYDKDEVDGRDCNCP 780
Db 721 PHLPNSGQEDFKDGIIGDADDDNDGVTDEKNCQLLFPNQADYDKDEVDGRDCNCP 780
QY 781 VYHNPAQIDTNNNGEGDACSVDIDGDDVFNERNDCPYVYNTDQRTDGDGVDHDCNCP 840
Db 781 VYHNPAQIDTNNNGEGDACSVDIDGDDVFNERNDCPYVYNTDQRTDGDGVDHDCNCP 840
QY 841 VHNPDQTDVNDLVGDQDNNEDIDDDGHQNNQDNCPIYISNANQADHSDGOGDADPDD 900
Db 841 VHNPDQTDVNDLVGDQDNNEDIDDDGHQNNQDNCPIYISNANQADHSDGOGDADPDD 900
QY 901 DNDGVPDRDNCRLVFNPDQEDLDGGRGDTCKDDFDNDNIPDIDDDVCPENNAISETDFR 960
Db 901 DNDGVPDRDNCRLVFNPDQEDLDGGRGDTCKDDFDNDNIPDIDDDVCPENNAISETDFR 960
QY 961 NFQMPVLPDKGTQIDPNWVTRHQKELVQVANSDPGLAVGFDEFGSDVSGTFVYNTDR 1020
Db 961 NFQMPVLPDKGTQIDPNWVTRHQKELVQVANSDPGLAVGFDEFGSDVSGTFVYNTDR 1020
QY 1021 DDDYAGFVFGYOSSRFYVVMKQVQTYWEDQPTFRAYGSGVSLKVVNSTGTGEHLRN 1080
Db 1021 DDDYAGFVFGYOSSRFYVVMKQVQTYWEDQPTFRAYGSGVSLKVVNSTGTGEHLRN 1080
QY 1081 ALWHTGNTPGQVRTLWHDPRNIGWKDYTAIRWHLTHRPKTYGIRVLVHEGKQVMAADSGPI 1140
```

```
Db 1081 ALWHTGNTPGQVRTLWHDPRNIGWKDYTAIRWHLTHRPKTYGIRVLVHEGKQVMAADSGPI 1140
QY 1141 YDOTYAGGRGLGIFVPSQEMVYFSDLYKECRDI 1172
Db 1141 YDOTYAGGRGLGIFVPSQEMVYFSDLYKECRDI 1172
RESULT 2
thrombospondin 2 precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 20-Aug-1999
C:Accession: A42587; A39851
R:Lahterty, C.D.; O'Rourke, K.; Wolf, F.W.; Katz, R.; Seldin, M.F.; Dixit, V.M.
J. Biol. Chem. 267, 3274-3281, 1992
A:Title: Characterization of mouse thrombospondin 2 sequence and expression during c
A:Reference number: A42587; MUID:92147683
A:Accession: A42587
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-1172 <LAH>
A:Cross-references: GB:L07803; GB:M87275; NID:g340421; PIDN:AAA53064.1; PID:g567241
A:Note: sequence extracted from NCBI backbone (NCBIP:81502)
R:Bornstein, P.; O'Rourke, K.; Wikstrom, K.; Wolf, F.W.; Katz, R.; Li, P.; Dixit, V.
J. Biol. Chem. 266, 12821-12824, 1991
A:Title: A second, expressed thrombospondin gene (thbs2) exists in the mouse genome.
A:Reference number: A39851; MUID:91302287
A:Accession: A39851
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-873 <BOR>
A:Cross-references: GB:M64866; NID:g201994; PIDN:AAA40432.1; PID:g201995
C:Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology;
E:Keywords: calcium binding; glycoprotein
F:319-377/Domain: von Willebrand factor type C repeat homology <VWC>
F:380-431/Domain: thrombospondin type 1 repeat homology <THR1>
F:436-492/Domain: thrombospondin type 1 repeat homology <THR2>
F:493-549/Domain: thrombospondin type 1 repeat homology <THR3>
F:553-588/Domain: EGF homology <EGF1>
F:652-691/Domain: EGF homology <EGF>
Query Match 89.6%; Score 5920; DB 2; Length 1172;
Best Local Similarity 88.5%; Pred. No. 0;
Matches 1036; Conservative 57; Mismatches 78; Indels 0; Gaps 0;
QY 1 MWRLVLLALWVWSTQAGHODKDTTDLFSISINIRKTIAGAKQRPDPGPVPAFYRVRF 60
Db 1 MLWALALLALGIGPRASAGDHVKDTSTDLFSISINIRKTIAGAKQRPDPGPVPAFYRVRF 60
QY 61 DYIPVNAADDLSKITKIMROKEGFFLTALQKODGKSRGTLALLEGPGISQRFVSNRP 120
Db 61 DYIPVNTDDLNRIVKLARRKEGFFLTALQKODRKSRTLLVLEGPGTSQRFVSNRP 120
QY 121 ADTLDLTWIDGTRHVSLEVDGLADSOQKNTVQVAGETYSLVHVCGLIDSFALDEFFY 180
Db 121 GDTLDLNTWVEGNQHTNFDVGLADSOQKNTVQVADTYSLVYGCGLIDSVTLEEFFY 180
QY 181 EHLQAEKSRMYVAKGASARESHFRGLLQNVHLVFNSEVEDILSKKCGQCGQGAENAISEN 240
Db 181 EQLVDRSMYVAKGASARESHFRGLLQNVHLVFNSEVEDILSKKCGQSGQAEVNTISEH 240
QY 241 TETLRLGHVHTVTEYVGPSSRRPEYCSCEBELGNMVOELSLHLVNLQPSNKRVSND 300
Db 241 TETLHLSPHITDLVVOGVEKAQEVCTHSCBELSNMNLSELGLHVMVNLKNLERVSSD 300
QY 301 NOFLWELIGGPPKTRNMSACHQDGRFFFAENETWVDSCTTCTCKFKTICHQITCPPATC 360
Db 301 NOFLLELIGGLKTRNMSACVQEGRIFAENETWVDSCTTCTCKFKTIVCHQITCPPATC 360
QY 361 ASPSFVEGECPCSLHSDVGEKSWPWAETQCSVTCSGSGTQQRGRSCDVTSTNTCLGPSI 420
Db 361 ASPSFVEGECPCSLHSDVGEKSWPWAETQCSVTCSGSGTQQRGRSCDVTSTNTCLGPSI 420
```

Db 361 ANPSFVEGCCPSCSHSADSDGSHSPWAETECSTVTCGSGTQGRSCDVTNTCLGPSI 420
QY 421 QTRACSLSKCDTRIQRDGGSHWSWSSCSVTGCVGNITRILCNLSPVPMQGGKCKGSG 480
Db 421 QTRFCSLGKCDTRIQRNGGSHWSWSSCSVTGCVGNITRILCNLSPVPMQGGKCKGSG 480
QY 481 RETKACOGACPTIDGRWSPHSPSACTVTCAGGIRTRVCNSPEPOYGKACVGDVQER 540
Db 481 RETKPCQORDPCPIDGRWSPHSPSACTVTCAGGIRSRVCNSPEPOYGKACVGDVTEH 540
QY 541 QMCNKRSCTPVDGCLSNPCFPAGCSPFPGDGSWSCGSPVGFNGTHCEDLDECAVPOI 600
Db 541 QMCNKRSCTPIDGCLSNPCFPAGCSEFPDGSWSCGSPVGFNGTHCEDLDECAVVDI 600
QY 601 CFSTSKVPCVNTQPGFHLCPCPRYRGNOFVGVGLEAAKTEKQVCEPENCKDKTHNCH 660
Db 601 CFSTNKAPRCVNTNPGFHLCPCPRYKGNQFPGVGLGLEDARTERKQVCEPENCKDKTHSCH 660
QY 661 KHAECIYLGHFSDPMYKCECOTGAGDGLICGEDSDLDGWPNNLVCATNATYHCKDKNC 720
Db 661 KHAECIYLGHFSDPMYKCECOTGAGDGLICGEDSDLDGWPNNLVCATNATYHCKDKNC 720
QY 721 PHLPNSQEDFDKDGIGDADCDNDNDGVTDERKDNQCLLPNPRQADYDKDEVDGRDCNCP 780
Db 721 PKLPNSQEDFDKDGIGDADCDNDNDGVSDEKDNQCLLPNPRQADYDKDEVDGRDCNCP 780
QY 781 YVHNPAQIDTDNNGEGDACSVDIDGDDVFNERNCPPYVNTDQDRTDGGVGHDCNCP 840
Db 781 YVHNPAQIDTDNNGEGDACSVDIDGDDVFNERNCPPYVNTDQDRTDGGVGHDCNCP 840
QY 841 VHPNDQTDVNDLVGQDCNNEDIDDDGHONNODNCPYISNANOADHRRDQGDGADCPD 900
Db 841 MHPNDQTDVNDLVGQDCNNEDIDDDGHONNODNCPYISNANOADHRRDQGDGADCPD 900
QY 901 DNDGVPDRDRNCLVFPNPDQEDLDGGRGDIKDDFDNDNIPDIDVCPENNAISETDF 960
Db 901 DNDGVPDRDRNCLVFPNPDQEDSDGGRGDIKDDFDNDNIPDIDVCPENNAISETDF 960
QY 961 NFQWPLDPKGTQIDPNWVIRHQKELVQTANSDFGIAVGFDFGVSDFSGTFYVNTDR 1020
Db 961 NFQWPLDPKGTQIDPNWVIRHQKELVQTANSDFGIAVGFDFGVSDFSGTFYVNTDR 1020
QY 1021 DDDYAGVFYQSSRFYVMVQVQTYWEDQTPRAYGSGYSLKVVNSTTGTGHELRN 1080
Db 1021 DDDYAGVFYQSSRFYVMVQVQTYWEDKPSRAYGSGYSLKVVNSTTGTGHELRN 1080
QY 1081 ALMHTGTNPGQVRLTHDPRNIGWKDYATYRWHLTHRPKTYIRVLVHEGKQVYADSGPI 1140
Db 1081 ALMHTGTNPGQVRLTHDPRNIGWKDYATYRWHLTHRPKTYIRVLVHEGKQVYADSGPI 1140
QY 1141 YDQTYAGRLGLFVFSQEMVYFSDLYKECRD 1171
Db 1141 YDQTYAGRLGLFVFSQEMVYFSDLYKECRD 1171

RESULT 3

A39804

thrombospondin precursor - chicken

C:Species: Gallus gallus (chicken)

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

R:Lawler, J.; Duquette, M.; Ferro, P.

J. Biol. Chem. 266, 8039-8043, 1991

A:Title: Cloning and sequencing of chicken thrombospondin.

A:Reference number: A39804; MUID:91217026

C:Accession: A39804

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1178 <LAW>

A:Cross-references: GB:M60853; NID:g212763; PID:AAA51437.1; PID:g212764

C:Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology; v

F:325-383/Domain: von Willebrand factor type C repeat homology <VWC>

F:386-437/Domain: thrombospondin type 1 repeat homology <THR1>

F:442-498/Domain: thrombospondin type 1 repeat homology <THR2>
F:499-555/Domain: thrombospondin type 1 repeat homology <THR3>
F:658-697/Domain: EGF homology <EGF>

Query Match 83.0%; Score 5481; DB 1; Length 1178;

Best Local Similarity 79.7%; Pred. No. 0;
Matches 934; Conservative 110; Mismatches 126; Indels 2; Gaps 2;

QY 1 MWRLVLLALWMPSTQAGHQDKDTTDFLSISINIRKKTIGAKQFRGDPGVPAYRFRVF 60

Db 7 LLWLAVFITLWSSDAQDDAKEENTFDLLQISINIRKKTIGAKLFRGDPDAIPAYRFR 66

QY 61 DYTPPNADSLKTIIMRQKEGFFLTAQLKQDKSRGTLALLEGPLSORQEIFVSNP 120

Db 67 DHIPPFKEPKLKVILIRQNEGFILSATLRQDRQSRGTLALLEGPLSQRQEIFVSNR 126

QY 121 ADPLDTIYIDGRHVRVSLVDGLADSQNKNTVQVAGETYSLVHVGDLDSFALDPFFY 180

Db 127 ANTLDLIYVWDFQNVISLVDGLADSQNKNTVQVAGETYSLVHVGDLDSFALDPFFY 186

QY 181 EHLQAEKSMYVAKGSAHSHFRLGLLQNVHLVENSVEDILSKKGGQGGQGAENAISEN 240

Db 187 EQLKAENSMYVAKGSTIRENHFRLGLLQNIHLIEDTSTEDVLRKKGQSRQSTEVNTINES 246

QY 241 TETLRGLPHVTTEYVGPSSERRPERCEELGNMVOELSLGLHVLVNQPSNLRKVSND 300

Db 247 TETLRGLPHVTTEYVGEKTEKAEFCDSCEELGTFTLTGRIYVNNADLNQKVSSE 306

QY 301 NQFLWELIGPPKT-RNMSACWQDGRFFFAENETWYVDSCTCTCKFKTKTCHQITCPPAT 359

Db 307 NQIMWELI-GPNKTLANQSVQWQDGRVADSESWIVDSCTCKTCQDSKIVCHOITCPPVS 365

QY 360 CASPSFVEGECPCSLHSDGEGSWPBAWTCVTCGSGTQGRSCDVTNTCLGPS 419

Db 366 CADPSFIEGECPCVCSHSDSEEGSWPBDWTKCVTCGSGTQGRSCDVTNTCLGPS 425

QY 420 IOTRACSLSKCDTRIQRDGGSHWSWSSCSVTGCVGNITRILCNLSPVPMQGGKCKGSG 479

Db 426 IOTRACSLSKCDTRIQRDGGSHWSWSSCSVTGCVGNITRILCNLSPVPMQGGKCKGSG 485

QY 480 GRETKACQAGAPCIDGRWSPWSPSACTVTCAGGIRTRVCNSPEPOYGKACVGDVQ 539

Db 486 GRETKACEKAPCPVNGQWSPWSPSACTVTCAGGIRTRVCNSPEPOYGKACVGDVQ 545

QY 540 RQMCNRSCTPVDGCLSNPCFPAGCSPFPGDGSWSCGSPVGFNGTHCEDLDECAV 599

Db 546 HDMCNKRDCEPIDGCLSNPCFPAGCSPFPGDGSWSCGSPVGFNGTHCEDLDECAV 605

QY 600 ICFTSKVPCVNTQPGFHLCPCPRYRGNOFVGVGLEAAKTEKQVCEPENCKDKTHNC 659

Db 606 VCFKVNQVHRCVNTNPGFHLCPCPRYKGNQFPGVGLGLEDARTERKQVCEPENCKDKTHSCH 665

QY 660 HKHAECIYLGHFSDPMYKCECOTGAGDGLICGEDSDLDGWPNNLVCATNATYHCKDK 719

Db 666 HKHAECIYLGHFSDPMYKCECOTGAGDGLICGEDSDLDGWPNNLVCATNATYHCKDK 725

QY 720 CPHLPNSQEDFDKDGIGDADCDNDNDGVTDERKDNQCLLPNPRQADYDKDEVDGRDCN 779

Db 726 CPHLPNSQEDFDKDGIGDADCDNDNDGVTDERKDNQCLLPNPRQADYDKDEVDGRDCN 785

QY 780 PYVHNPAQIDTDNNGEGDACSVDIDGDDVFNERNCPPYVNTDQDRTDGGVGHDCNCP 839

Db 786 PYVHNPAQIDTDNNGEGDACSVDIDGDDVFNERNCPPYVNTDQDRTDGGVGHDCNCP 845

QY 840 LVHNPQTDVNDLVGQDCNNEDIDDDGHONNODNCPYISNANOADHRRDQGDGADCPD 899

Db 846 LVHNPQTDVNDLVGQDCNNEDIDDDGHONNODNCPYISNANOADHRRDQGDGADCPD 905

QY 900 DDNDGVPDRDRNCLVFPNPDQEDLDGGRGDIKDDFDNDNIPDIDVCPENNAISETDF 959

Db 906 DDNDGVPDRDRNCLVFPNPDQEDLDGGRGDIKDDFDNDNIPDIDVCPENNAISETDF 965

QY	960	RNFQWFLDPKGGTTQIDPNWVIRHQKELVQTANSPGIATVGDFGVSDFSGTFYYNTD	1019
Db	966	RKFQWFLDPKGGTAQIDPNWVIRHQKELVQTANSPDGIATVGDFGVSDFSGTFYYNTD	1025
QY	1020	RDDDYAGVFVGYSOSSRFYVVMKKQVQTQYVWEDQPTFRAGYSGVSLKVVNSTTGTGGEHLR	1079
Db	1026	RDDDYAGVFVGYSOSSRFYVVMKKQVQTQYVWEDKTRAYGYSVSLKVVNSTTGTGGEHLR	1085
QY	1080	NALMHTGNTGCGVRTLWHDPRTNGWKDYATYARWHLITHRPKGTGYIRVLVHEGKQVMADSGP	1139
Db	1086	NALMHTGNTGCGVRTLWHDPKNTGNWKDYATYARWHLITHRPKGTGLIKVLVEGKQVMVDSGP	1145
QY	1140	IYDTYAGGRLGLGVFSQEMVYFSDLKYECD	1171
Db	1146	IYDTFAGGRLGLGVFSQEMVYFSDLKYECD	1177
RESULT	4		
TSUP1			
		thrombospondin 1 precursor - human	
		C:Species: Homo sapiens (man)	
		C:Date: 23-Aug-1987 #sequence_revision 03-Aug-1995 #text_change 17-Nov-2000	
		C:Accession: A26155; A34274; A30140; A25812; A05172; A42927	
		R:Lawler, J.; Hynes, R.O.	
		J. Cell Biol. 103, 1635-1648, 1986	
		A:Title: The structure of human thrombospondin, an adhesive glycoprotein with	
		A:Reference number: A26155; MUID:87057617	
		A:Accession: A26155	
		A:Molecule type: mRNA	
		A:Residues: 1-1170 <LAW>	
		A:Cross-references: GB:X04665; NID:g37137; PIDN:CAA28370.1; PID:g37138	
		A:Note: parts of this sequence, including the amino end of the mature prote	
		R:Lahterty, C.D.; Gierman, T.M.; Dixit, V.M.	
		J. Biol. Chem. 264, 11222-11227, 1989	
		A:Title: Characterization of the promoter region of the human thrombospondin	
		A:Reference number: A34274; MUID:89291870	
		A:Accession: A34274	
		A:Molecule type: DNA	
		A:Residues: 1-166 <LAH>	
		A:Cross-references: GB:J04835	
		R:Hennessy, S.W.; Frazier, B.A.; Kim, D.D.; Deckwerth, T.L.; Baumgartel, D.N.	
		J. Cell Biol. 108, 729-736, 1989	
		A:Title: Complete thrombospondin mRNA sequence includes potential regulatory	
		A:Reference number: A30140; MUID:89139590	
		A:Accession: A30140	
		A:Molecule type: mRNA	
		A:Residues: 1-83; 'A', '85-522', 'A', '524-1170' <HEN>	
		A:Cross-references: EMBL:X14787; NID:g37464; PIDN:CAA32889.1; PID:g37465	
		A:Note: parts of this sequence, including the amino end of the mature prote	
		R:Kobayashi, S.; Eden-McCuttchan, F.; Framson, P.; Bornstein, P.	
		Biochemistry 25, 8418-8425, 1986	
		A:Title: Partial amino acid sequence of human thrombospondin as determined t	
		A:Reference number: A25812; MUID:87157592	
		A:Accession: A25812	
		A:Molecule type: mRNA	
		A:Residues: 1-83; 'A', '85-397' <KOB>	
		A:Cross-references: GB:M25631; NID:g538353; PIDN:AAA36741.1; PID:g538354	
		R:Dixit, V.M.; Hennessy, S.W.; Grant, G.A.; Rotwein, P.; Frazier, W.A.	
		Proc. Natl. Acad. Sci. U.S.A. 83, 5449-5453, 1986	
		A:Reference number: A05172; MUID:86287276	
		A:Accession: A05172	
		A:Molecule type: mRNA	
		A:Residues: 1-83; 'A', '85-374', 'RC' <DIX>	
		A:Cross-references: GB:M4326; NID:g340005; PIDN:AAA61237.1; PID:g553801	
		A:Note: parts of this sequence, including the amino end of the mature prote	
		R:Sun, X.; Skorstengaard, K.; Mosher, D.F.	
		J. Cell Biol. 118, 693-701, 1992	
		A:Title: Disulfides modulate RGD-inhibitable cell adhesive activity of throm	
		A:Reference number: A42927; MUID:92348511	
		A:Accession: A42927	
		A:Molecule type: protein	
		A:Residues: 987-1003 <SUN>	
		A:Note: Cys-992 is shown to have a free sulfhydryl	

C:Genetics:

A:Gene: GDB:THBS1; TSP1; TSP

A:Cross-References: GDB:120436; OMIM:188060

A:Map position: 15q15-15q15

A:Introns: 23/1

A>Note: the list of introns may be incomplete

C:Complex: homotrimer, disulfide linked

C:Function:

A:Description: participates in cell migration and adhesion, and in platelet aggregation

C:Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology

C:Keywords: beta-hydroxyasparagine; calcium binding; cell adhesion; glycoprotein; trypsin

F:1-18/Domain: signal sequence #status predicted <SIG>

F:19-1170/Product: thrombospondin 1 #status predicted <MAT>

F:317-375/Domain: von Willebrand factor type C repeat homology <VWC>

F:378-423/Domain: thrombospondin type 1 repeat homology <THR>

F:434-490/Domain: thrombospondin type 1 repeat homology <THR>

F:491-547/Domain: thrombospondin type 1 repeat homology <THR2>

F:551-586/Domain: EGF homology <EGF1>

F:587-689/Domain: EGF homology <EGF2>

F:650-689/Domain: EGF homology <EGF1>

F:926-928/Region: cell attachment (R-G-D) motif

F:171-232/Disulfide bonds: #status predicted

F:248,360,708,1067/Binding site: carboxydrate (Asn) (covalent) #status predicted

F:270,274/Disulfide bonds: interchain #status predicted

F:610/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted

F:1051/Binding site: carbohydrate (Asn) (covalent) #status absent

Query Match	64.0%;	Score 4228;	DB 1;	Length 1170;
Best Local Similarity	61.6%;	Pred. No. 4.5e-250;		
Matches 725;	Conservative 170;	Mismatches 267;	Indels 14;	Gaps 6;
Qy 1	MYWRL-VLLALWMPSTQAGHQDKD-TTFDLFSISNNRKTIGAKQFRGPDGPVPAYRFV 58			
Db 3	LAWGLGVLFMHVCGNRRIPESGGDINSVFDFELTGAARKSGRRLYKGPDPSSPAFRLE 62			
Qy 59	RFDYTPPVNADLSKTKTKMRQKEGFLLTAQLKODGKSRGFTLLALEGGLSORQFEIVSN 118			
Db 63	DANLIPVPDDKFDQLVDVARTKEGFLLLASLQMKKTGRTGLLALERKDHSGQVFSVYN 122			
Qy 119	GPADTLDLYTWDGTRHWVSLEVDGLADSQWKNVTVOVAGETYSLRHVGCDDLDGFALDEP 178			
Db 123	GRAGTLDLSLTVOGRKHVVSVEALLATGQWKSITLFVQEDRAQLYIDCEKMEALDVP 182			
Qy 179	F---YEHLOAKRSMTYKVAAGSRESHFRGLLQNVHLVFENSVEDILSKKQCGQOGABIN 235			
Db 183	IQSVFTRDLASLARLAKGVND-RFQGVQNVRFVFGTTPEDILRNKGCSSSTSVLLT 241			
Qy 236	AISENTETLRGPHVTEVGPSPERRPVCERSCEBLGNMVQELSLHVLVNPSENLK 295			
Db 242	L--DNNVNGSSPAINTNYGHKTDLQATCGISDCELSMWLELRGLRTIVTTLQDSIR 299			
Qy 296	RYSNDNQFLWELIGGPPKTRNKSACQWDGFFAENETWYVDSCTCTCKFKTKTICHQITC 355			
Db 300	KYTEENKELANELRRPP-----LCYHNGVQYRNNEWTVDSCTECHQNSVTICKKVCSC 353			
Qy 356	PPATCASPSFVGECCPSCLHSVDGSGEGHSFWAEWTQCVSTCGSGTQORGRSCDVTNWC 415			
Db 354	PIMPCSNATVPDGECCPRCPMSDASDADGWSPSWEWTSCSTSCGNGIQGRGSCDSLNNRC 413			
Qy 416	LGPSLTQTRACLSKCDTRTRDQGGWSHSPWSSCVTCGVGNITRIRLCHNSPVPQMGKN 475			
Db 414	EGSSVQTRTCHTQEBCKRFQDGGWSHSPWSSCVTCGGGVTRIRLCHNSPSPQMGKP 473			
Qy 476	CKSGRETAKCAGACPIDGRNSPWSPSACTVTCAGGIRERTRVCSNPBPQYGGKACVG 535			
Db 474	CEGEARETKACKKADKAPINGWGWPMDTCSVTCGGGVQKRSLCNPNPTPQFGKGCDCVG 533			
Qy 536	DVQERQMNKRCSPVDGCLSNPCFPCAQCSSFPDGSWSGSCSPVGLFGNTHCDDLDECA 595			
Db 534	DVTENQICNKQDCPIDGCLSNPCFPAGVKCTSYPDGSWKACGACPPGYSNGIQCTYDVECK 593			
Qy 596	LVPDTCFSTSKVPRCVNTQPGFHLCPCPPRYRGQNPQVGVGLEAAKTEKQVCEPNPKCK 655			

Db 594 EVPDACFNHNGEHRCEHTDPGYNCLPCPPRETGSQPPGGQVEHATANKQVCKPRNPCTDS 653
Qy 656 THNCHKHAEICYLGHSDPMYKCECQTAGDGLIGCBSDLDGPNLNLVCATNATYHC 715
Db 654 THDCNNAKCNLYGHYSDPMYKCECKPGVAGNGIICGEDTDLGPNENLVCVANATYHC 713
Qy 716 IKNCPLHNSGQEDFKDGDIGDADCDNDDNDGVTDEKNCOLLFPNROADYDKBVGDR 775
Db 714 KDCNPLHNSGQEDYDKDGDIGDADCDNDDNDKIPDRNCPPHYNPAQYDYDRDDVGDR 773
Qy 776 CDMCPYVHNPAQIDTNNEGDACSVDIDGDDVFNERNDCPPYVNTDQRTDGDGVDGHC 835
Db 774 CDMCPYVHNPAQIDTNNEGDACAADIDGDLNERNDCQYVYVNDQRTDMDGVGDQC 833
Qy 836 DMCPLVNPQDQVNDLVGDCQDNEDIDGCHQNNQDCPPIISANQADHRRDQGDGDA 895
Db 834 DMCPLVNPQDQVNDLVGDCQDNEDIDGCHQNNQDCPPIISANQADHRRDQGDGDA 893
Qy 896 CDPDDNDGVPDRDNCRLVFNPDQEDLDGDRGDKCDKDFDNDNDPDIIDVCPENNAIS 955
Db 894 CDHDDNDGVPDRDNCRLVFNPDQEDLDGDRGDKCDKDFDNDNDPDIIDVCPENNAIS 953
Qy 956 ETDNRNFMVPLDKTQTDIDPNMVIHQKGLVQVANSDPGIAVGDFGVSDFSGTFY 1015
Db 954 ETDNRNFMVPLDKTQTDIDPNMVIHQKGLVQVANSDPGIAVGDFGVSDFSGTFY 1013
Qy 1016 VNTDRDDYAGVFGVQSSRFYVWVKQVQTYWEDQTPRAYGSGVSLKVVNSTGTG 1075
Db 1014 INTERDDYAGVFGVQSSRFYVWVKQVQTYWEDQTPRAYGSGVSLKVVNSTGTG 1073
Qy 1076 EHLRNALWHTGNTPGQVRLWHDPRNGWKTAYRWHLTHRPKTYIRVLVHEGQVMA 1135
Db 1074 EHLRNALWHTGNTPGQVRLWHDPRNGWKTAYRWHLTHRPKTYIRVLVHEGQVMA 1133
Qy 1136 DSGPIVDQTYAGGRLGLFVSQEMVYFSDLYKECRD 1171
Db 1134 DSGPIVDQTYAGGRLGLFVSQEMVYFSDLYKECRD 1169

RESULT 5
A40558
Thrombospondin 1 precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 05-Jun-1992 #sequence revision 05-Jun-1992 #text_change 20-Aug-1999
C:Accession: A40558; A37905; B42587; S68787
R:Lawler, J.; Duquette, M.; Ferro, P.; Copeland, N.G.; Gilbert, D.J.; Jenkins, N.A.
Genomics 11, 587-600, 1991
A:Title: Characterization of the murine thrombospondin gene.
A:Reference number: A40558; MUID:92128941
A:Accession: A40558
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1170 <LAW>
A:Cross-references: GB:M62449; GB:M62450; GB:M62451; GB:M62452; GB:M62453; GB:M62454; GB:M62465; GB:M62466; GB:M62467; GB:M62468; GB:M62469; GB:M62470; NID:9511867; PIDN:AAA5061
R:Bornstein, P.; Alf, D.; Devaratnu, S.; Framson, P.; Li, P.
J. Biol. Chem. 265, 16691-16698, 1990
A:Title: Characterization of the mouse thrombospondin gene and evaluation of the role of
A:Reference number: A37905; MUID:90375546
A:Accession: A37905
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-490 <BOR>
A:Cross-references: GB:J05605; GB:J05606; NID:9201991; PIDN:AAA40431.1; PID:9554390
R:Laherty, C.D.; O'Rourke, K.; Wolf, F.W.; Katz, R.P.; Seldin, M.F.; Dixit, V.M.
J. Biol. Chem. 267, 3274-3281, 1992
A:Title: Characterization of mouse thrombospondin 2 sequence and expression during cell
A:Reference number: A42587; MUID:92147683
A:Accession: A42587
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1152, 'p', 1154-1170 <LAH>
A:Cross-references: GB:M87276

A:Note: sequence extracted from NCBI backbone (NCBIP:81501)
R:Chen, H.; Aeschlimann, D.; Nowlen, J.; Mosher, D.F.
FEBS Lett. 387, 36-41, 1996
A:Title: Expression and initial characterization of recombinant mouse thrombospondin
A:Reference number: S68787; MUID:96234006
A:Accession: S68787
A:Molecule type: protein
A:Residues: 19-26, 'X', 28-37 <CHE>
C:Complex: homotrimer, disulfide linked
C:Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-1170/Product: thrombospondin 1 #status predicted <MAT>
F:317-375/Domain: von Willebrand factor type C repeat homology <VWC>
F:378-423/Domain: thrombospondin type 1 repeat homology <THR1>
F:434-490/Domain: thrombospondin type 1 repeat homology <THR2>
F:491-547/Domain: thrombospondin type 1 repeat homology <THR3>
F:551-586/Domain: EGF homology <EGF>
F:248,360,708,1067/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 63.4%; Score 4189.5; DB 2; Length 1170;
Best Local Similarity 61.8%; Pred. No. 1e-247;
Matches 723; Conservative 162; Mismatches 272; Indels 13; Gaps 5;

Qy 6 VLLALWVWSTQAGHDKDT-TDFLFSISNINRKTIGAKQFRGPDGPGVAYRFRFDYIP 64
Db 9 VLLFLHMGCSNRLPESGGDNGVDFIDELIGGARGRRLKVGQDLSSFAFRLENANLIP 68
Qy 65 PYNADDLSKITIMROKEGFFLTAQLKQDGKSRGTLLEGGPGLSQRFQEIFSVNGPADTL 124
Db 69 AVPDCKFQDLDDAVWADKGFIFLASLRQMKTRGTLIAVERKDNTGQIFSVVSNKAGTL 128
Qy 125 DLTWYWDGTRHYVSLDVGADLSQWKNVTQVAGETYSLVHGCDLDSFALDEP---FYE 181
Db 129 DLSLSLPGKQVYVSEALLATGQWKSITLFVQEDRAQLYIDCKMESAEFLDPIOSIF 188
Qy 182 HLOAEKSRMYVAKGSARESHFGLLONVHLVFNESVEDILSKKGCQGGCAEINASENT 241
Db 189 RLASVARLVRKAGDVND-NFGVLQNVRFVGTTFEDILRNKGCSSNTNLLTL--DNN 245
Qy 242 ETLRLGPHVTYVGPSPERRPERCERSEELGNMVOELSLHLVNLQPSNLRKVSNDN 301
Db 246 VVNGSSPAIRNTYIGHTKDLQALICGLSDELSSWLELKLGLRTIVTLQDSIRKVTEN 305
Qy 302 QFLWELIGPPKTRNMSACWDGRFAENETWVVDSCCTCTCKKFKTICHQITCPATCA 361
Db 306 RELVSELKRPP-----LCFHNGVQYKNNEEWTVDSCTECHCONSVTICKKVCPIMPCS 359
Qy 362 SPSEFVEGECPCSLHSDGEGWSPWAEWTQCSVTCSGTQGRGRSCDVT-SNTCLGPSIQ 421
Db 360 NATVPDECCPRGWSDSADDDGWSPWSEWTSCSATNGIQGRSCDSLNNRCESSVQ 419
Qy 422 TRACLSKCDTRIRQDGGHSHSPWSSCSVTGCVGNITRILNCSVPVQMGKNKKGSGR 481
Db 420 TRTCHIQECDKRFQDGGHSHSPWSSCSVTGCDGVITRILNCSVPVQMGKNKKGSGR 479
Qy 482 ETKACGACPCPDGRWSPSPSACTVTCAGGIRTRVCNCSPEPQYGGKACVGDVQERQ 541
Db 480 ETKACKKADACPTNGGWPSPMDICSVTCGGVQRSLCNPNTPQFGKDCVGDVTENQ 539
Qy 542 MCKNRSCPVDGCLSNPCFPGAOCSPFDGSGWSCGPGVFLNGTHCEDLDECALVPDIC 601
Db 540 VCNKQDCPIDGCLSNPCFAGAKTSTPDGSKWCKACPPGYSNGNGIOCKVDDECKEVPDAC 599
Qy 602 FSTSKVPRCVNTQPGFCHLPCPPRYRGNQPVGVGLEAAKTEKQVCEPENPCDKTINCHK 661
Db 600 FHNHGERCKNTDPGYNCLPCPPRFTGSPFGRGVEHAMANKQVCKPRNPCTDGTDCNK 659
Qy 662 HABCYILGHFSDPMYKCECQTQTAGDGLICGEDSDLDGPNLNLVCATNATYICIDNCP 721
Db 660 NAKCNLYGHYSDPMYKCECKPGVAGNGIICGEDTDLGPNENLVCVANATYHCKKDNCP 719

Matches 360; Conservative 78; Mismatches 183; Indels 69; Gaps 11;

QY 537 VQERQMC---NKRSCPVGDCLSNCPFGAQS-----SPDGSWSGCGSPVGLNGTHCED 590
Db 263 IMECVCGFHEQRS-----HCSNPFCRGVDCMEVEYEP--GYRCGPCPPGLQNGTHCSD 316
QY 591 LDECALVPDTCFSTSKVPRCVNTOPGFHCLPCPPRYRGNOQPVGVGLEAAKTEKQVCE--- 647
Db 317 INECAHA-DPCFPSS---CINTMPGFHCEACPRGYGTQVSGVGDYARASKQVCNDID 372
QY 648 -----PENPCDKTKHN-CHKHAEC 665
Db 373 ECNDGNGGCDPNSICTNTVTGSKGCGPCLGFLGNOSQGCVPARTCHSPAHSPCHIIAHC 432
QY 666 IYLGHFSDPMYKCECOTGYAGDGLICGEDSDLDGWPMLNLVCAATNATYHICIKNCPLPN 725
Db 433 LF---ERNAGVSCQCNVWAGNVGCPDIDGYPDOALPCMDN-NKHCKQDNCLLTPN 488
QY 726 SGQEDFDKDGIGDADDDDDNDGVTDEKNCQLLFNPRQADYDKDEVDGDCDCNCPYVHNP 785
Db 489 SGQEDADNDGVCQDQDDADGDKIKNVEDNCRFLFPNKDQNSDTSFSGDACDCNCPYVNN 548
QY 786 AQIDTDNNGEGDACSVDIDGDDVFNERNDCPYVNTDQRTDGGVGDHDCNCPLVHNP 845
Db 549 DQKDTGNGEGDADNDVDGDIPIGLDNCPKVPNPLOTDRDEGVGDACDSCPEMSNPT 608
QY 846 QTDVDNDLVGDCDNNEDIDDDGHQNNQDNCPIYSNANQADHNRDGDGDCDPCDDNDGV 905
Db 609 QTDADSLVGVDCVTDNEDSGDGHQDTKNCPLPNSQLSDNDGLGDCDGDNDGI 668
QY 906 PD-----DRDNCRLVFNPDQEDLDGGRGDKODFDNDNIPDIDVCPENNAISETDFRN 961
Db 669 PDYVPPGDCNRLVFNPNQKSDGNGVDCEDDFDNDVVDPLDVCPEAEVTLTDFRA 728
QY 962 QVMPVLPDKGTTDIDPNWVLRHOGKELVQANSDPGTAIGFDFGSGVDFSTGYVNTFRD 1021
Db 729 YQTVLDPEDGDAQIDPNWVLRHOGKELVQANSDPGTAIGFDFGSGVDFSTGYVNTFRD 788
QY 1022 DDVAGFVGVGSSRFVVMKQVOTYQWEDQTPTRAVGSGVSLKVVNTGTGGEHLNA 1081
Db 789 DDVAGFLFSTQDSGRFVVMKQVOTYQWEDQTPTRAVGSGVSLKVVNTGTGGEHLNA 848
QY 1082 LWHGTGTPGVRRILWHDPRNIGKDYATYRHLTHRPKTYIRVLVHEGKQVMADSGPIY 1141
Db 849 LWHGTGTPGVRRILWHDPRNIGKDYATYRHLTHRPKTYIRVLVHEGKQVMADSGPIY 908
QY 1142 DQYVAGRLGLFVFSQEMVYFSDLKYECD 1171
Db 909 DTSMRGRLGVFCFSQENIIWSNLQVRCND 938
RESULT 8
A46016
thrombospondin 3 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jan-2000
C:Accession: A46016; A44124; I55398; S68788; S72433
R:Bornstein, P.; Devaryalu, S.; Edelhoff, S.; Distech, C.M.
Genomics 15, 607-613, 1993
A:Title: Isolation and characterization of the mouse thrombospondin 3 (Thbs3) gene.
A:Reference number: A46016; MUID:93224149
A:Accession: A46016
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-956 <BOR>
A:Cross-references: GB:I04302; NID:g202200; PIDN:AAA0497.1; PID:g202201
A:Note: sequence extracted from NCBI backbone (NCBIN:129415, NCBI:129416)
R:Vos, H.L.; Devaryalu, S.; de Vries, Y.; Bornstein, P.
J. Biol. Chem. 267, 12192-12196, 1992
A:Title: Thrombospondin 3 (Thbs3), a new member of the thrombospondin gene family.
A:Reference number: A44124; MUID:92291102
A:Accession: A44124
A>Status: preliminary

A:Molecule type: DNA
A:Residues: 517-956 <VOS>
A:Cross-references: GB:M86620; NID:g201987; PIDN:AAA0430.1; PID:g201989
A:Experimental source: BALB/c
A:Note: sequence extracted from NCBI backbone (NCBIN:106634, NCBIN:106636, NCBIN:106638)
R:Qabar, A.N.; Lin, Z.; Wolf, F.W.; O'Shea, K.S.; Lawler, J.; Dixit, V.M.
J. Biol. Chem. 269, 1262-1269, 1994
A:Title: Thrombospondin 3 is a developmentally regulated heparin binding protein.
A:Reference number: I55398; MUID:94117438
A:Accession: I55398
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-432, 'F', '434-562', 'D', '564-719', 'E', '721-870', 'W', '872-956' <RES>
A:Cross-references: GB:L24434; NID:g402718; PIDN:AAA0433.1; PID:g402719
R:Chen, H.; Aeschlimann, D.; Nowien, J.; Mosher, D.F.
FEBS Lett. 387, 36-41, 1996
A:Title: Expression and initial characterization of recombinant mouse thrombospondin
A:Reference number: S68787; MUID:96234006
A:Accession: S68788
A:Molecule type: protein
A:Residues: 'X', '23-28', 'X', '30', 'X', '32' <CHE>
R:Collins, M.; Bornstein, P.
Nucleic Acids Res. 24, 3661-3669, 1996
A:Title: SP1-binding elements, within the common metaxin-thrombospondin 3 intergenic
A:Reference number: S72433; MUID:97025352
A:Accession: S72433
A:Molecule type: DNA
A:Residues: 1-26 <COL>
A:Cross-references: EMBL:U66257; NID:g1575552; PIDN:AAC52819.1; PID:g1575554
C:Genetics:
A:Gene: Thbs-3
C:Complex: homopentamer, disulfide linked
C:Superfamily: thrombospondin 3; EGF homology
C:Keywords: calcium binding; glycoprotein; homopentamer
F:374-412/Domain: EGF homology <EGF>
F:310,407,644,937/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 29.3%; Score 1936.5; DB 1; Length 956;
Best Local Similarity 52.0%; Pred. No. 2.4e-110;
Matches 359; Conservative 81; Mismatches 181; Indels 69; Gaps 11;

QY 537 VQERQMC---NKRSCPVGDCLSNCPFGAQS-----SPDGSWSGCGSPVGLNGTHCED 590
Db 263 IMECVCGFHEQRS-----HCSNPFCRGVDCMEVEYEP--GYRCGPCPPGLQNGTHCSD 316
QY 591 LDECALVPDTCFSTSKVPRCVNTOPGFHCLPCPPRYRGNOQPVGVGLEAAKTEKQVCE--- 647
Db 317 INECAHA-DPCFPSS---CINTMPGFHCEACPRGYGTQVSGVGDYARASKQVCNDID 372
QY 648 -----PENPCDKTKHN-CHKHAEC 665
Db 373 ECNDGNGGCDPNSICTNTVTGSKGCGPCLGFLGNOSQGCVPARTCHSPAHSPCHIIAHC 432
QY 666 IYLGHFSDPMYKCECOTGYAGDGLICGEDSDLDGWPMLNLVCAATNATYHICIKNCPLPN 725
Db 433 LF---ERNAGVSCQCNVWAGNVGCPDIDGYPDOALPCMDN-NKHCKQDNCLLTPN 488
QY 726 SGQEDFDKDGIGDADDDDDNDGVTDEKNCQLLFNPRQADYDKDEVDGDCDCNCPYVHNP 785
Db 489 SGQEDADNDGVCQDQDDADGDKIKNVEDNCRFLFPNKDQNSDTSFSGDACDCNCPYVNN 548
QY 786 AQIDTDNNGEGDACSVDIDGDDVFNERNDCPYVNTDQRTDGGVGDHDCNCPLVHNP 845
Db 549 DQKDTGNGEGDADNDVDGDIPIGLDNCPKVPNPLOTDRDEGVGDACDSCPEMSNPT 608
QY 846 QTDVDNDLVGDCDNNEDIDDDGHQNNQDNCPIYSNANQADHNRDGDGDCDPCDDNDGV 905
Db 609 QTDADSLVGVDCVTDNEDSGDGHQDTKNCPLPNSQLSDNDGLGDCDGDNDGI 668
QY 906 PD-----DRDNCRLVFNPDQEDLDGGRGDKODFDNDNIPDIDVCPENNAISETDFRN 961
Db 669 PDYVPPGDCNRLVFNPNQKSDGNGVDCEDDFDNDVVDPLDVCPEAEVTLTDFRA 728

```
Db 669 PDYIPPGDNCRLVNPKNOKSDGNGVGDVCEDEDFDNDVAVDPLDVCPEAGVTLTDFRA 728
QY 962 EQWVPLDPKGTQTIDPNWVIRHQRELVTQANSDPGIAVGDFGSDVDFSGTFYVNTDRE 1021
Db 729 YITVILDPDGGDAIDPNWVLLNOGMEIVQTMNSDPGLAVGYTAFNGVDVDFEGTFHYNTYTD 788
QY 1022 DDYAGFVFGYQSSRFYVVMQVQTYWEDQPTRAYGYSGLKVMNSTTGTGTEHLRNA 1081
Db 789 DDYAGFLFSQDSGRFYVVMQVQTYWEDQPTRAYGYSGLKVMNSTTGTGTEHLRNA 848
QY 1082 LHTGNTGQVFTLWHDPRNIGKDYTAYRWHLTHRPKTYIRVLVHKGQVMADSGPIY 1141
Db 849 LHTGHTPDQVRLWLTDPNVLGRDKTSYRWLLHRPQVGYIRVKLYEGPQLVADSGVII 908
QY 1142 DOTYAGGLGLFVFSQEMVYFSDLYECRD 1171
Db 909 DYSMRGGLGVFCFSQENIINSLQYRCND 938

RESULT 9
A45441
thrombospondin 4 - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jan-2000:
C:Accession: A45441
R:Lawler, J.; Whittaker, M.; Whittaker, C.A.; Adams, J.C.; McHenry, K.; DeSimone, D.W.
J. Cell Biol. 120, 1059-1067, 1993
A:Title: Identification and characterization of thrombospondin-4, a new member of the th
A:Reference number: A45441; MUID:93163109
A:Accession: A45441
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-955 <LAW>
A:Cross-references: GB:219091; NID:q288777; PIDN:CAA79518.1; PID:q288778
A:Note: sequence extracted from NCBI backbone (NCBIN:124858, NCBI:124860)
C:Superfamily: thrombospondin 3; EGF homology
F:285-319/Domain: EGF homology <EGF>

Query Match 28.8%; Score 1899; DB 2; Length 955;
Best Local Similarity 51.9%; Pred. No. 4.6e-108; Mismatches 184; Indels 60; Gaps 7;
Matches 348; Conservative 79;

QY 549 PVDGCLSNPCFPGAQCSFFPDGWSGSCPVGFLNGTHCEDLDPCALVPDIFCSTSKVP 608
Db 281 PKPRCDATSCFRGVRCID-TEGFGCGCPCEYTGNGVICTDVBDCRLNP--CFLG---V 334
QY 609 RCVNTQPGFHLCPPRYRGNQPVGVGLEAAKTEKQV----- 645
Db 335 RCINTSPGFKCSCPPGYTGSTIQIGINFARQKQVQCTDTNECENGRNGGCTSNLGIN 394
QY 646 -----CEPENPCDKTHNCKHAECTYLGHFSDPMYKCECQGY 684
Db 395 TMSFRGCGKPGYVGDQIKCKPKESKHGQNPCHASAQC---SEKDGVDVTCFSGV 451
QY 685 AGDGLIGEDSDLDGWPNLNLVCATNATYHCIDKNCPLNPSGQBDKDGIDGACDDDD 744
Db 452 AGNGYLCGKDTIDGYDEALPCPDK---NCKKNCVVPVNSGQEDTDKDNIGDACCDEA 508
QY 745 DNDGVTDEKNCQLLFNPRQADYKDEVDGRCDNCPYVHNPAQIDTDNNGSGDACSVDID 804
Db 509 DGDGLNEQDNCVLAANTDKNSQDIFGDACDNCRLTLNNDQRDITDNGRGDACDDMD 568
QY 805 GDDVFNERNDCPYVYNTDQRTDGDGVDHDCNCPLVHNPDTQVNDLVGDGDCDNNEDI 864
Db 569 GDGIKNIIDNCORVPNVQDKDGDGVGIDCDSCPDIIINPQSDIDNDLVGSDCCTNODS 628
QY 865 DDDGHQNDQNCPTYSRANQADHDGQDGDACDPDDDDGVDPD---DRDNCRLVFNPDQ 920
Db 629 DGDGHODSTDCPTVINSQJLDTKDGIDGDECDDDDDNDGIDPDTVPYPCDNCCLVNPNGQ 688
QY 921 EDLDGSDGRGDKDQDFDNDNTPIDDVCPENNAISETDFRNFMVPLDPKGTQIDPNWY 980
Db 980 -----
```

```
Db 689 EDDNNDGVGDVCEADFDDQDVTIDRIDVCPENAEITLTDFRAYQTVVLDPEGDAQIDPNWI 748
QY 981 IRHQKELVQVANSDPGIAVGDFGSDVDFSGTFYVNTDREDDDYAGFVFGYQSSRFYV 1040
Db 749 VLNOGMEIVQTMNSDPGLAVGYTAFNGVDFEGTFHYNTMTDDDDYAGFIFGYQDSSFYV 808
QY 1041 MKQVQTYWEDQPTRAYGYSGLKVMNSTTGTGTEHLNALWHTGNTGQVRTLWHDPR 1100
Db 809 MKQVQTYWEDQPTRAYGYSGLKVMNSTTGTGTEHLNALWHTGNTGQVRTLWHDPR 868
QY 1101 NGWKDYTYAYRWHLTHRPKTYIRVLVHKGQVMADSGPIYDQTYAGGRGLGFVFSQEMV 1160
Db 869 NYGWDKYSYRWFLQHRPQVGYIRARFVEGTDELVDGVTVDTRMGRGLGVFCFSQENI 928
QY 1161 YFSDLKYECRD 1171
Db 929 IWSNLKYRCND 938

RESULT 10
TSHUP4
thrombospondin 4 precursor - human
C:Species: Homo sapiens (man)
C:Date: 06-Jan-1995 #sequence_revision 11-Aug-1995 #text_change 13-Aug-1999
C:Accession: A55710; S36069
R:Lawler, J.; McHenry, K.; Duquette, M.; Derick, L.
J. Biol. Chem. 270, 2809-2814, 1995
A:Title: Characterization of human thrombospondin-4.
A:Reference number: A55710; MUID:95155352
A:Accession: A55710
A:Molecule type: mRNA
A:Residues: 1-961 <LAW>
A:Cross-references: EMBL:Z19585; NID:g311625; PIDN:CAA79635.1; PID:g311626
A:Note: authors translated the codon GTG for residue 616 as Ser
C:Genetics:
A:Gene: GDB:THBS4
A:Cross-references: GDB:463011; OMIM:600715
A:Map position: lq21-lq23
A:Complex: homotrimer, disulfide linked
C:Function:
A:Description: participates in cell migration and adhesion, and in platelet aggregati
C:Superfamily: thrombospondin 3; EGF homology
C:Keywords: beta-hydroxyasparagine; beta-hydroxyaspartic acid; calcium binding; cell
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-961/Product: thrombospondin 4 #status predicted <MAT>
F:290-324/Domain: EGF homology <EGF1>
F:330-362/Domain: EGF homology <EGF>
F:562-584/Region: cell attachment (R-G-D) motif
F:303/Modified site: erythro-beta-hydroxyaspartic acid (Asp) #status predicted
F:343/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted
F:612,941/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 28.6%; Score 1890; DB 1; Length 961;
Best Local Similarity 53.1%; Pred. No. 1.6e-107;
Matches 355; Conservative 71; Mismatches 185; Indels 58; Gaps 9;

QY 549 PVDGCLSNPCFPGAQCSFFPDGWSGSCPVGFLNGTHCEDLDPCALVPDIFCSTSKVP 608
Db 286 PPRCDNSNPCFRGVOCTDSRQD-FQCGPCPEGYTGNGITCIDVDECKYHP--CYPG---V 339
QY 609 RCVNTQPGFHLCPPRYRGNQPVGVGLEAAKTEKQV-----EPEN----- 650
Db 340 HCINLSPGFRCDACPVGTPGMVQGVGISFAKSNKQVCTDIDECRNGACVPNSICVNTLG 399
QY 651 -----PK-----DKTHNCKHAECTYLGHFSDPMYKCECQGYTAG 686
Db 400 SYRGCPCKPGYTGDIQIRGCKVERNCNPELNPSCSYNAQCI---EERQGDVTCVCGVGAG 456
QY 687 DGLICGEDSDLDGWPNLNLVCATNATYHCIDKNCPLNPSGQEDKDGIDGACDDDDDN 746
Db 457 DGYICGDKVDIDSYDPDEELPCSR---NCKKDNCKYVNPNSGQEDADRDGIGDACCDEADG 513
```


[illegible]

SAI137PM	;ZCT
AATLVAIA	SSINOC
65;	MISMACTLES
188;	TIDEITS
230;	GAPS
Z/;	

99 TLLALEGPGLSQR-----QFEIVSNCPADTLDTYWDGT-----RHVVS 138
 897 TCSATCGEGLSRERSRCRRGSCWEDDASQTRRCVNGPCEHSYLT-WSFWTTCETCSFDS 955

QY 282 GLHLVNPSENKRVSNQFLWELIGGP-PKTRNMSACWDGRFFAENETWVVDSCTT 340
Db 345 -----EGFPAPQTGDPAA-----EWSPMWCVCSSTC 370
QY 341 -----CTCKKFTICHQITCPATCASPSFVEGECPCSLHSDVDEEGWSPWAEW 390
Db 371 GEGWQTRTRCVSSYST-----QCSGPLREORLCNNSAVCPVHG--ANDEWSPW 418
QY 391 TQCSVTCSGTQOGRSC---DVTNLTCLGPSIQTRACSLSKCDTRIRQDGGWSHWSPWS 447
Db 419 SLASSTCGRFRDTRTCRPPQFGGNPCGEPKQTKFCNIALCPGR-AVDGNWNEWSSWS 477
QY 448 SCSVTCGVGNITRIRLCLNSVPQMGKNCKGSGRETRACOGAPCPIDGRWSPWSPHSACT 507
Db 478 ACSASCSQGRQRTRECNG--PSYGGAECCGHVWYETRDCLQOCPCVDGKWQAWASWGSCS 535
QY 508 VTCAGGIRERTRVCNSPEPOYGGKACVGYOERQMCNKRSCP--DGCLSNPCFPFGAQCS 565
Db 536 VTCGAGSORRERVCSG--PFFGGAACOGPODEYRQCGRCPPEPHEIC-DEDNFGAVIWK 592
QY 566 SFPDGSWSCGSCPVGFLNGT-----HCEDLDE 593
Db 593 ETPAGEVAAVRCP-----RNATGLILRRCE-LDE 620

Search completed: August 9, 2002, 10:02:15
Job time: 171 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2002, 10:01:40 ; Search time 15.18 seconds
(without alignments)
2989.415 Million cell updates/sec

Title: US-09-822-682-2
Perfect score: 6605
Sequence: 1 MWRLVLLALWVPSQAGH.....FVFSQEMVYFDLKYECRDI 1172

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5
105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6570	99.5	1172	1 TSP2_HUMAN	P35442 homo sapien
2	6068	91.9	1170	1 TSP2_BOVIN	Q95116 bos taurus
3	5920	89.6	1172	1 TSP2_MOUSE	Q03350 mus musculus
4	5481	83.0	1178	1 TSP2_CHICK	P35440 gallus gall
5	4228	64.0	1170	1 TSP1_HUMAN	P07996 homo sapien
6	4228	63.9	1170	1 TSP1_BOVIN	Q28178 bos taurus
7	4189.5	63.4	1170	1 TSP1_MOUSE	P35441 mus musculus
8	4189	63.4	1173	1 TSP1_XENLA	P35448 xenopus lae
9	1993	30.2	757	1 COMP_HUMAN	P49747 homo sapien
10	1968	29.8	755	1 COMP_RAT	P35444 rattus norv
11	1946.5	29.5	956	1 TSP3_HUMAN	P49746 homo sapien
12	1936.5	29.3	956	1 TSP3_MOUSE	Q05895 mus musculus
13	1899	28.8	955	1 TSP4_XENLA	Q06441 xenopus lae
14	1890	28.6	961	1 TSP4_HUMAN	P35443 homo sapien
15	1881	28.5	980	1 TSP4_RAT	P49744 rattus norv
16	1480	22.4	439	1 COMP_BOVIN	P35445 bos taurus
17	439.5	6.7	437	1 PROP_MOUSE	P11680 mus musculus
18	424	6.4	1384	1 BAI1_HUMAN	Q14514 homo sapien
19	418.5	6.3	810	1 BAI1_RAT	Q62919 rattus norv
20	418	6.3	1522	1 BAI3_HUMAN	Q60242 homo sapien
21	416	6.3	469	1 PROP_HUMAN	P27918 homo sapien
22	415	6.3	1074	1 SMSA_HUMAN	Q13591 homo sapien
23	407.5	6.2	470	1 PROP_CAVPO	Q64181 cavia porce
24	391	5.9	1077	1 SMSA_MOUSE	Q62217 mus musculus
25	390	5.9	1093	1 SMSB_MOUSE	Q60519 mus musculus
26	389.5	5.9	816	1 NEL2_MOUSE	Q61220 mus musculus
27	388.5	5.9	810	1 NEL1_HUMAN	Q92832 homo sapien
28	384.5	5.8	1572	1 BAI2_HUMAN	Q60241 homo sapien
29	379.5	5.7	816	1 NEL2_RAT	Q62918 rattus norv
30	366.5	5.5	816	1 NEL_CHICK	Q90827 gallus gall
31	361.5	5.5	816	1 NEL2_HUMAN	Q99435 homo sapien
32	343.5	5.2	2531	1 NTC1_MOUSE	Q01705 mus musculus
33	340.5	5.2	2531	1 NTC1_RAT	Q07008 rattus norv

ALIGNMENTS

RESULT 1

ID	TSP2_HUMAN	STANDARD;	PRT;	1172 AA.
AC	P35442;			
DT	01-JUN-1994 (Rel. 29, Created)			
DT	01-JUN-1994 (Rel. 29, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Thrombospondin 2 precursor.			
GN	THBS2 OR TSP2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=94010892; PubMed=8406456;			
RA	Labell T.L., Byers P.H.;			
RT	"Sequence and characterization of the complete human thrombospondin 2			
RT	cdNA: potential regulatory role for the 3' untranslated region.";			
RL	Genomics 17:225-229(1993).			
RN	[2]			
RP	SEQUENCE OF 560-1172 FROM N.A.			
RC	TISSUE= Fibroblast;			
RX	MEDLINE=92217961; PubMed=1559694;			
RA	Labell T.L., McGookey Milewicz D.J., Distche C.M., Byers P.H.;			
RT	"Thrombospondin II: partial cdna sequence, chromosome location, and			
RT	expression of a second member of the thrombospondin gene family in			
RT	humans.";			
RL	Genomics 12:421-429(1992).			
CC	-!- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND			
CC	CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,			
CC	LAMININ AND TYPE V COLLAGEN.			
CC	-!- SUBUNIT: HOMOPRIMER; DISULFIDE-LINKED.			
CC	-!- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.			
CC	-!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.			
CC	-!- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.			
CC	-!- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.			
CC	-!- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL; L12350; AAA03703.1; -			
DR	EMBL; M81339; -; NOT_ANNOTATED_CDS.			
DR	PIR; A42173; A42173.			
DR	HSP; P35555; 1EMN.			
DR	MIM; 188061; -			
DR	InterPro; IPR000561; EGF-like.			
DR	InterPro; IPR001881; EGF_Ca.			
DR	InterPro; IPR001791; Laminin_G.			
DR	InterPro; IPR000884; TSPl.			

```
DR InterPro: IPR003129; TSPN:
DR InterPro: IPR001007; VFWC:
DR InterPro: IPR003367; tsp_3:
DR Pfam: PF00008; EGF_2:
DR Pfam: PF02210; TSPN; 1:
DR Pfam: PF00090; tsp_1; 3:
DR Pfam: PF02412; tsp_3; 9:
DR Pfam: PF00093; VWC; 1:
DR SMART: SM00181; EGF; 3:
DR SMART: SM00209; TSP1; 3:
DR SMART: SM00210; TSPN; 1:
DR SMART: SM00214; VWC; 1:
DR PROSITE: PS00022; EGF_1; FALSE_NEG.
DR PROSITE: PS01186; EGF_2; 1:
DR PROSITE: PS00092; TSP1; 3:
DR PROSITE: PS01208; VFWC; 1:
KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
EGF-like domain; Signal.
FT SIGNAL 1 18
FT CHAIN 19 1172
FT DOMAIN 19 232
FT DOMAIN 318 375
FT DOMAIN 381 432
FT DOMAIN 437 493
FT DOMAIN 494 548
FT DOMAIN 549 589
FT DOMAIN 590 647
FT DOMAIN 648 692
FT DOMAIN 725 760
FT DOMAIN 761 783
FT DOMAIN 784 819
FT DOMAIN 820 842
FT DOMAIN 843 880
FT DOMAIN 881 916
FT DOMAIN 917 952
FT DOMAIN 953 1172
FT SITE 928 930
FT DISULFID 266
FT DISULFID 270
FT DISULFID 553
FT DISULFID 558
FT DISULFID 577
FT DISULFID 594
FT DISULFID 601
FT DISULFID 622
FT DISULFID 652
FT DISULFID 659
FT DISULFID 680
FT CARBOHYD 151
FT CARBOHYD 316
FT CARBOHYD 330
FT CARBOHYD 457
FT CARBOHYD 584
FT CARBOHYD 710
FT CARBOHYD 1069
SQ SEQUENCE 1172 AA; 129955 MW; 2AC7BB230E44C6F5 CRC64;

Query Match 99.5%; Score 6570; DB 1; Length 1172;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1167; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 MYRLVLLALWYWPSTQAGHQDKTTFDLFSISINRKTIGAKQFRGPDGVPAYRFVRF 60
DB 1 MYRLVLLALWYWPSTQAGHQDKTTFDLFSISINRKTIGAKQFRGPDGVPAYRFVRF 60
QY 61 DYIPPNADDLSKITKIMRQKEGFFLTALQKODGSRGTLTALLEGPGLSQROFEIVSNGP 120
DB 61 DYIPPNADDLSKITKIMRQKEGFFLTALQKODGSRGTLTALLEGPGLSQROFEIVSNGP 120
QY 121 ADTLDTLTYWIDGTRHVSLEDEGLADSQWKNVTQVAGETYSLHVGCDDLIDSFALDEPFY 180
DB 121 ADTLDTLTYWIDGTRHVSLEDEGLADSQWKNVTQVAGETYSLHVGCDDLIGPVALDEPFY 180
```

RESULT 2

TSP2_BOVIN

ID TSP2_BOVIN

STANDARD:

PRT: 1170 AA.

AC O95116; Q28180;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Thrombospondin 2 precursor (Corticotropin-induced secreted protein)
DE (CISP).
GN THS2 OR TSP2 OR TSP-2.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID:9913;
RN [1]
RA SEQUENCE FROM N.A.
RA Danik M., Chinn A., Lafuillade M., Keramidas M., Aquesse-Germon S.,
RA Penhoat A., Chen H., Mosher D., Chambaz E.M., Feige J.J.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-522 FROM N.A.
RX MEDLINE-96331130; PubMed-8698834;
RA Lafuillade B., Pellerin S., Keramidas M., Danik M., Chambaz E.M.,
RA Feige J.J.;
RT "Opposite regulation of thrombospondin-1 and corticotropin-induced
RT secreted protein/thrombospondin-2 expression by adrenocorticotrophic
RT hormone in adrenocortical cells";
RL J. Cell. Physiol. 167:164-172(1996).
RN [3]
RP SEQUENCE OF 318-831 FROM N.A.
RC TISSUE-Aortic endothelium;
RA Zafar R.S., Moll Y.D., Womack J.F., Walz D.A.;
RT "Cloning and sequencing of bovine thrombospondin stimulatory effect of
RT TGF-beta";
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
CC CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
CC LAMININ AND TYPE V COLLAGEN.
CC -1- SUBUNIT: HOMODIMER; DISULFIDE-LINKED.
CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC -1- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X96540; CAA65385.1; -;
CC EMBL; X87620; CAA60952.1; -;
CC InterPro: IPR000561; EGF-like.
CC InterPro: IPR001881; EGF-Ca.
CC InterPro: IPR001791; Laminin_G.
CC InterPro: IPR000884; TSP1.
CC InterPro: IPR003129; TSPN.
CC InterPro: IPR001007; VWFC.
CC InterPro: IPR003367; tsp_3.
CC Pfam: PF00008; EGF; 2.
CC Pfam: PF02210; TSPN; 1.
CC Pfam: PF00090; tsp_1; 3.
CC Pfam: PF02412; tsp_3; 9.
CC Pfam: PF00093; vwc; 1.
CC SMART: SM00181; EGF; 3.
CC SMART: SM00209; TSP1; 3.
CC SMART: SM00210; TSPN; 1.
CC SMART: SM00214; VWC; 1.
CC PROSITE: PS00022; EGF_1; FALSE_NEG.
CC PROSITE: PS01186; EGF_2; 1.
CC PROSITE: PS50092; TSP1; 3.
CC PROSITE: PS01208; VWFC; 1.

KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 1170 THROMBOSPONDIN 2.
FT DOMAIN 19 232 HEPARIN-BINDING (POTENTIAL).
FT DOMAIN 318 375 VWFC.
FT DOMAIN 379 430 TSP TYPE-1 1.
FT DOMAIN 435 491 TSP TYPE-1 2.
FT DOMAIN 492 546 TSP TYPE-1 3.
FT DOMAIN 547 587 EGF-LIKE 1.
FT DOMAIN 588 645 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 646 690 EGF-LIKE 3.
FT DOMAIN 723 758 TSP TYPE-3 1.
FT DOMAIN 759 781 TSP TYPE-3 2.
FT DOMAIN 782 817 TSP TYPE-3 3.
FT DOMAIN 818 840 TSP TYPE-3 4.
FT DOMAIN 841 878 TSP TYPE-3 5.
FT DOMAIN 879 914 TSP TYPE-3 6.
FT DOMAIN 915 950 TSP TYPE-3 7.
FT DOMAIN 951 1170 C-TERMINAL.
FT SITE 926 928 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 266 266 INTERCHAIN (PROBABLE).
FT DISULFID 270 270 INTERCHAIN (PROBABLE).
FT DISULFID 551 562 BY SIMILARITY.
FT DISULFID 556 572 BY SIMILARITY.
FT DISULFID 575 586 BY SIMILARITY.
FT DISULFID 592 608 BY SIMILARITY.
FT DISULFID 599 617 BY SIMILARITY.
FT DISULFID 620 644 BY SIMILARITY.
FT DISULFID 650 663 BY SIMILARITY.
FT DISULFID 657 676 BY SIMILARITY.
FT DISULFID 678 689 BY SIMILARITY.
FT CARBOHYD 151 151 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 316 316 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 455 455 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 582 582 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 708 708 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 936 936 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1067 1067 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 535 535 A -> V (IN REF. 3).
FT CONFLICT 748 748 S -> T (IN REF. 3).
SQ SEQUENCE 1170 AA; 129862 MW; 9CF1BF55B89A051 CRC64;

Query Match 91.9%; Score 6068; DB 1; Length 1170;
Best Local Similarity 90.0%; Pred. No. 0;
Matches 1055; Conservative 62; Mismatches 53; Indels 2; Gaps 1;

QY 1 MWRLVLLALWVWPSTQAGHQDKDTTDLFSISINIKRTIGAKQFRGPDGVPAYRVRF 60
DB 1 MLWPLLLALWVWPSTQAGHQDKDTTDLFSISINIKRTIGAKQFRGPDGVPAYRVRF 60
QY 61 DYTPPVNADLSKITKIMRQEGFFLTAQLKQDGKSGRTLLALRGGLSGLSRQFEIVSNGP 120
DB 61 DYTPPVSAEHLGRITAEAMRRKEGFFLTASMKQDRSGRTLLALRGGLSGLSRQFEIVSNGP 120
QY 121 ADTLDITYWIDGTRHVVSLEDVGLADSQWKNVTVQVAGETYSLHVGCDLDSFALDEPFF 180
DB 121 ADTLDITYWIDGTRHVVSLEDVGLADSQWKNVTVQVAGETYSLHVGCDLDSFALDEPFF 180
QY 181 EHLQAEKSRMYVAKGSAESHFRGLQNVHLVPENSVEDILSKKGGCOGGAENALSEN 240
DB 181 EHLQAEKSRMYVAKGSAESHFRGLQNVHLVPENSVEDILSKKGGCOGGAENALSEN 240
QY 241 TETLRGLPHVTVTEYVGPSSRRRPEVCERSCEELGNMYQELSGHLVNLQPSSENKRYVND 300
DB 241 TETLRGLPHVTVTEYVGPSSRRRPEVCERSCEELGNMYQELSGHLVNLQPSSENKRYVND 300
QY 301 NQFLWELIGGPPKTRNNSACWQGRFFAENETWVVDSCCTCTCKKFKTICHQITCPTATC 360
DB 301 NQFLWELIGGPPKTRNNSACWQGRFFAENETWVVDSCCTCTCKKFKTICHQITCPTATC 360

FT	DOMAIN	843	880	TSP TYPE-3 5.	
FT	DOMAIN	881	916	TSP TYPE-3 6.	
FT	DOMAIN	917	952	TSP TYPE-3 7.	
FT	DOMAIN	953	1172	C-TERMINAL.	
FT	SITE	928	930	CELL ATTACHMENT SITE (POTENTIAL).	
FT	DISULFID	266	266	INTERCHAIN (PROBABLE).	
FT	DISULFID	270	270	INTERCHAIN (PROBABLE).	
FT	DISULFID	553	564	BY SIMILARITY.	
FT	DISULFID	558	574	BY SIMILARITY.	
FT	DISULFID	577	588	BY SIMILARITY.	
FT	DISULFID	594	610	BY SIMILARITY.	
FT	DISULFID	601	619	BY SIMILARITY.	
FT	DISULFID	622	646	BY SIMILARITY.	
FT	DISULFID	652	665	BY SIMILARITY.	
FT	DISULFID	659	678	BY SIMILARITY.	
FT	DISULFID	680	691	BY SIMILARITY.	
FT	CARBOHYD	151	151	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	316	316	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	330	330	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	457	457	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	584	584	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	710	710	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	1069	1069	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	SEQUENCE	1172	AA; 129911 MW; 7CE8E4E8599822AB CRC64;		
Query Match					89.6%; Score 5920; DB 1; Length 1172;
Best Local Similarity					88.5%; Pred. No. 0;
Matches 1036; Conservative					57; Mismatches 78; Indels 0; Gaps 0;
QY	1	MVRLVLLALWVPSTQAGHDKDFTDFLFSINIRKTIKAGKQFRGPDGPAYRFRF	60		
DB	1	MLMALALLALGIGPRASGDHVKDTSDFLFSINIRKTIKAGKQFRGPDGPAYRFRF	60		
QY	61	DIYPPVADLKSITIMRKGFFFTLAOLKQDKSRGTLALLEGCLSORQEIYVNSGP	120		
DB	61	DIYPPVNTDNLRIVLARKKEGFFFTLAOLKQDKSRGTLALLEGCLSORQEIYVNSGP	120		
QY	121	ADTLDTYWDGRRHVSLEVDGLADSQMKNVTQVAGETYSLVHVGCDLIDSALDEPY	180		
DB	121	GDTLDLNLYWVEGQHTNFLEVDGLADSQMKNVTQVADSTYSLVVGCDLIDSALDEPY	180		
QY	181	EHLQAKSRMYAKGASRSHFGLLQNVHLVFNESVEDILSKKCGQOGAENAISEN	240		
DB	181	EQLEVDKSRMYAKGASRSHFGLLQNVHLVFNESVEDILSKKSCQHSGGAENVITISEH	240		
QY	241	TETLRGLPHVTYEVGSSRPERPEVCSCEELGNMVQELSLGLHLVLPNSENLKRVSND	300		
DB	241	TETLRGLSPHITDLVYQVGEKAQEVCTHSCSELSNMNELSGLHVMVNLKSLRVSND	300		
QY	301	NQFLWELIGGPPKTRMNSACWQGRFAENETWVVDSCCTCTCKKFKTCHQITCPTATC	360		
DB	301	NOFLELIGGLPLTRMNSACVQEGRIFAENETWVVDSCCTCTCKKFKTCHQITCPTATC	360		
QY	361	ASPSFVEGECPCSLHSVDGEGSWAETQCSVTGCGSQOGRSCDVTSNTCLGPSI	420		
DB	361	ANPSFVEGECPCSHASDSDEGSWAEWTECSVTGCGSQOGRSCDVTSNTCLGPSI	420		
QY	421	QTRACSLSKCDTRIRDDGSHWSPSSCSVTGCGVGNITRILCNLSPVPMGSKNCKGSG	480		
DB	421	QTRTCSLGLKCDTRIRQNGGSHWSPSSCSVTGCGVGNITRILCNLSPVPMGSKNCKGSG	480		
QY	481	RETKACQAPCDIDGRWSPSPWSACTVTCAGGIRTRVCNPSPEYQYGGKACGVQDER	540		
DB	481	RETKPCQORDCPTIDGRWSPSPWSACTVTCAGGIRRSRVCNPSPEYQYGGKACGVQTEH	540		
QY	541	QMCNKRKCPVDCGLNCPGPAQCSFPGDSWSCGCPVGFNGTHCEDLDCAVVDI	600		
DB	541	QMCNKRKCPIDGCLNCPFPGAKNCNFPDGSWSCGCPVGFNGTHCEDLDCAVVTDI	600		
QY	601	CFSTKVPKRCVNTQPGFCHLCPKPPRYRGNOQVGVLEAAKTRQVCEPENCKDKTHNCH	660		
DB	601	CFSTNAPKRCVNTNPGFCHLCPKPPRYKGNQPGVGLDARTEKQVCEPENCKDKTHSCH	660		

QY	661	KHAEIYILGHFSDPMYKCBQCTGYAGDGLICGSDSLDGRPNLNLVCATNATYHCICDNC	720		
DB	661	KHAEIYILGHFSDPMYKCECQIGYAGDGLICGSDSLDGRPNLNLVCATNATYHCICDNC	720		
QY	721	PHLPNSGQEDFDKDGIGDADCDNDNDGVDYDEKNCQLLFPNQADYDKDEVGDRDNCNP	780		
DB	721	PKLPNSGQEDFDKDGIGDADCDNDNDGVDYDEKNCQLLFPNQADYDKDEVGDRDNCNP	780		
QY	781	YVHNPAQIDTDNNGEGDAGSVIDGDDVFNERNDCPYVYNTDQDRTDGDGCGVGRDNCNPL	840		
DB	781	YVHNPAQIDTDNNGEGDAGSVIDGDDVFNERNDCPYVYNTDQDRTDGDGCGVGRDNCNPL	840		
QY	841	VHNPQDQTDVNDLVGQDCDNNEDIDDDGHQNNQDNCPIYSNANOADHRDQGGDADCPDD	900		
DB	841	MHNPQDQDQNDLVGQDCDNNEDIDDDGHQNNQDNCPIYSNANOADHRDQGGDADCPDD	900		
QY	901	DNDGVDDRDNCRLVFNPOEDLDGGRGDIKDDDFDNDNIPDIDDDVCPENNAISETDFR	960		
DB	901	DNDGVDDRDNCRLVFNPOEDLDGGRGDIKDDDFDNDNIPDIDDDVCPENNAISETDFR	960		
QY	961	NFQMVPLDPKGTQIDPNVWIRHQGKELVQTANSDFGIAVGFDEFGSVDFSGTFYVNTDR	1020		
DB	961	NFQMVPLDPKGTQIDPNVWIRHQGKELVQTANSDFGIAVGFDEFGSVDFSGTFYVNTDR	1020		
QY	1021	DDYAGVFGYQSSSRFYVVMKQVTTQTYWEDQTPRAYGYSGVSLKVVNSTTGTGHELRN	1080		
DB	1021	DDYAGVFGYQSSSRFYVVMKQVTTQTYWEDQTPRAYGYSGVSLKVVNSTTGTGHELRN	1080		
QY	1081	ALWHTGNTQGVRTLWHDPRNTGKDYATRWHLTHRPKTGYIRLVHHEGKQVYADSGPI	1140		
DB	1081	ALWHTGNTQGVRTLWHDPRNTGKDYATRWHLTHRPKTGYIRLVHHEGKQVYADSGPI	1140		
QY	1141	YDQTYAGGRGLGFVFSQEMVYFSDLYKECRD	1171		
DB	1141	YDQTYAGGRGLGFVFSQEMVYFSDLYKECRD	1171		
RESULT 4					
TSP2_CHICK					
ID	TSP2_CHICK	STANDARD;	PRT;	1178 AA.	
AC	P35440;				
DT	01-JUN-1994 (Rel. 29, Created)				
DT	01-JUN-1994 (Rel. 29, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Thrombospondin 2 precursor.				
GN	THBS2 OR TSP2.				
OS	Gallus gallus (Chicken).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;				
OC	Gallus.				
OX	NCBI_TaxID=9031;				
RN	[1]				
SEQUENCE FROM N.A.					
RX	MEDLINE=91217026; PubMed=2022631;				
RA	Lawler J., Duquette M., Ferro P.;				
RT	"Cloning and sequencing of chicken thrombospondin."				
RL	J. Biol. Chem. 266:8039-8043(1991).				
CC	-/- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN, LAMININ AND TYPE V COLLAGEN.				
CC	-/- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.				
CC	-/- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.				
CC	-/- SIMILARITY: CONTAINS 1 VWFC DOMAIN.				
CC	-/- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.				
CC	-/- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.				
CC	-/- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial				

Oy 1140 IYDTYAGGRGLGLFVSEQWYFSDLYKXCRD 1171
 Db 1146 IYDTTFAAGRGLGLFVSEQWYFSDLYKXCRD 1177

RESULT 5
 TSP1_HUMAN STANDARD; PRT; 1170 AA.
 AC P07996;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Thrombospondin 1 precursor.
 GN THS1 OR TSP1 OR TSP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Endothelial cells;
 RX MEDLINE=87057617; PubMed=2430973;
 RA Lawler J., Hynes R.O.;
 RT "The structure of human thrombospondin, an adhesive glycoprotein with
 RT multiple calcium-binding sites and homologies with several different
 RT proteins.";
 RL J. Cell Biol. 103:1635-1648(1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89139590; PubMed=2918029;
 RA Hennessy S.W., Frazier B.A., Kim D.D., Deckwerth T.L.,
 RA Baumgartel D.M., Rotwein P., Frazier W.A.;
 RT "Complete thrombospondin mRNA sequence includes potential regulatory
 RT sites in the 3' untranslated region.";
 RL J. Cell Biol. 108:729-736(1989).
 RN [3]
 RP SEQUENCE OF 1-397 FROM N.A.
 RX MEDLINE=87157592; PubMed=3030396;
 RA Kobayashi S., Eden-McCutchan F., Framson P., Bornstein P.;
 RA Dixit V.M., Hennessy S.W., Grant G.A., Rotwein P., Frazier W.A.;
 RT "Partial amino acid sequence of human thrombospondin as determined by
 RT analysis of cDNA clones: homology to malarial circumsporozoite
 RT proteins.";
 RL Biochemistry 25:8418-8425(1986).
 RN [4]
 RP SEQUENCE OF 1-374 FROM N.A.
 RX MEDLINE=86287276; PubMed=3461443;
 RA Dixit V.M., Hennessy S.W., Grant G.A., Rotwein P., Frazier W.A.;
 RT "Characterization of a cDNA encoding the heparin and collagen binding
 RT domains of human thrombospondin.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:5449-5453(1986).
 RN [5]
 RP SEQUENCE OF 1-166 FROM N.A.
 RX MEDLINE=89291870; PubMed=2544587;
 RA Laherty C.D., Gierman T.M., Dixit V.M.;
 RT "Characterization of the promoter region of the human thrombospondin
 RT gene. DNA sequences within the first intron increase transcription.";
 RL J. Biol. Chem. 264:11222-11227(1989).
 RN [6]
 RP SEQUENCE OF 1028-1170 FROM N.A.
 RA la Fleur M., Jobin C., Gauthier J., Kreis C.G.;
 RT Submitted (XXX-1992) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP CARBOHYDRATE-LINKAGE SITES W-385; S-394; W-438; W-441; T-450; W-498
 RP AND T-507.
 RC TISSUE=platelet;
 RX PubMed=11067851;
 RA Hofsteenge J., Huwiler K.G., Macek B., Hess D., Lawler J.,
 RA Mosher D.F., Peter-Katalinic J.;
 RT "C-mannosylation and O-fucosylation of the thrombospondin type 1
 RT module.";
 RL J. Biol. Chem. 276:6485-6498(2001).
 CC -1- FUNCTION: Adhesive glycoprotein that mediates cell-to-cell and

cell-to-matrix interactions. Can bind to fibrinogen, fibronectin,
 laminin, type V collagen and integrins alpha-V/beta-1, alpha-
 V/beta-3 and alpha-IIB/beta-3.
 -1- SUBUNIT: Homotrimer; disulfide-linked.
 -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
 -1- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
 -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
 -1- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; M25631; AAA36741.1; -;
 EMBL; X04665; CAA28370.1; -;
 EMBL; X14787; CAA32889.1; -;
 EMBL; J04835; AAA61178.1; -;
 EMBL; M99425; AAB59366.1; -;
 PIR; A05172; A05172;
 PIR; A25812; A25812;
 PIR; A26155; A26155;
 PIR; A30140; A30140;
 PIR; A34274; A34274;
 HSP; P35555; IEMN.
 GlycoSuiteDB; P07996; -;
 MM; 188060; -;
 InterPro; IPR000561; EGF-like.
 InterPro; IPR001881; EGF-Ca.
 InterPro; IPR001791; Laminin_G.
 InterPro; IPR000884; TSP1.
 InterPro; IPR003129; TSPN.
 InterPro; IPR001007; VWFC.
 InterPro; IPR003367; tsp_3.
 Pfam; PF00008; EGF; 2.
 Pfam; PF02210; TSPN; 1.
 Pfam; PF00090; tsp_1; 3.
 Pfam; PF02412; tsp_3; 8.
 Pfam; PF00093; vwc; 1.
 SMART; SM00181; EGF; 3.
 SMART; SM00209; TSP1; 3.
 SMART; SM00210; TSPN; 1.
 SMART; SM00214; VWC; 1.
 PROSITE; PS00022; EGF_1; FALSE_NEG.
 PROSITE; PS01186; EGF_2; 1.
 PROSITE; PS50092; TSP1; 3.
 PROSITE; PS01208; VWFC; 1.
 KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
 KW EGF-like domain; Signal.
 FT SIGNAL 1 18
 FT CHAIN 19 1170
 FT DOMAIN 19 232
 FT DOMAIN 316 373
 FT DOMAIN 379 430
 FT DOMAIN 435 491
 FT DOMAIN 492 548
 FT DOMAIN 549 587
 FT DOMAIN 588 645
 FT DOMAIN 646 690
 FT DOMAIN 723 758
 FT DOMAIN 759 781
 FT DOMAIN 782 817
 FT DOMAIN 818 840
 FT DOMAIN 841 878
 FT DOMAIN 879 914
 FT DOMAIN 915 950
 FT DOMAIN 951 1170
 FT SITE 926 928
 FT DISULFID 270 270
 THROMBOSPONDIN 1.
 HEPARIN-BINDING (POTENTIAL).
 VWFC.
 TSP TYPE-1 1.
 TSP TYPE-1 2.
 TSP TYPE-1 3.
 EGF-LIKE 1.
 EGF-LIKE 2.
 EGF-LIKE 3.
 TSP TYPE-3 1.
 TSP TYPE-3 2.
 TSP TYPE-3 3.
 TSP TYPE-3 4.
 TSP TYPE-3 5.
 TSP TYPE-3 6.
 TSP TYPE-3 7.
 C-TERMINAL.
 CELL ATTACHMENT SITE (POTENTIAL).
 INTERCHAIN (PROBABLE).

```
FT DISULFID 274 INTERCHAIN (PROBABLE).
FT DISULFID 551 BY SIMILARITY.
FT DISULFID 572 BY SIMILARITY.
FT DISULFID 575 BY SIMILARITY.
FT DISULFID 592 BY SIMILARITY.
FT DISULFID 599 BY SIMILARITY.
FT DISULFID 620 BY SIMILARITY.
FT DISULFID 650 BY SIMILARITY.
FT DISULFID 657 BY SIMILARITY.
FT DISULFID 678 BY SIMILARITY.
FT CARBOHYD 248 N-LINKED (GLCNAC. . .).
FT CARBOHYD 360 C-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 385 C-LINKED (MAN. . .) (POTENTIAL).
FT CARBOHYD 394 /FTIG-CAR_000205.
FT CARBOHYD 438 /FTIG-CAR_000206.
FT CARBOHYD 441 /FTIG-CAR_000207.
FT CARBOHYD 450 /FTIG-CAR_000208.
FT CARBOHYD 458 /FTIG-CAR_000209.
FT CARBOHYD 498 /FTIG-CAR_000210.
FT CARBOHYD 507 /FTIG-CAR_000211.
FT CARBOHYD 708 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1067 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 84 T -> A (IN REF. 2, 3 AND 4).
FT CONFLICT 523 T -> A (IN REF. 2).
SQ SEQUENCE 1170 AA; 129412 MW; 69B3E5AE3A395E CRC64;
```

Query Match 64.0%; Score 4228; DB 1; Length 1170;
Best Local Similarity 61.6%; Pred. No. 5.2e-257; Indels 14; Gaps 6;
Matches 725; Conservative 170; Mismatches 267;

```
QY 1 MVRL-VLLALWMPSTQAGHQDKD-TTFDLFSISINRKTIGAKQFRGPDGVPAYRFV 58
DB 3 LAGLVFLMHVCGTNRIPESGDSNVDFELTCAARKSGRRLVKGPDSPSPAFRIE 62
QY 59 RFDYIPVNAVDSLTKIMRQEGFLTAQLKODKSGRTLLALSGPGLSQRQFIVSN 118
DB 63 DANLIPVPDDKQDLVDVARTKEGFLLLASLQMKRTKRTLLALERKDHSGQVFSVSN 122
QY 119 GPADTLDTYIDGTHRVYSLDVLGADSLQKNTVVOVAGETVSLHVGCDLISFALDEP 178
DB 123 GKAGTLDLSLTVOGKHVSVSEALLATGOWKSITLFEVDRAQLYIDCEKMEAEIDVP 182
QY 179 F---YEHLQAEKSMYVAKGSAHREHFRGLLQNVHLVFNSEVBIILSKGCGQGGAEIN 235
DB 183 IQSVFTRLASIALRIAKGVND-NFQGVQLQNVRFVFTTPBIDLNRKGSSTSVLLT 241
QY 236 AISENTELRGLPHVTVTEYVGPSSRRPVCERSCELGNNVQELSLGLHVLVNPQSENK 295
DB 242 L--DNNVNGVSSPAIRTNVIGHKTKLQALQICGISCDELSSWLELRLGRTVITVLQDSIR 299
QY 296 RVSDNQFWEILGGPKTRNMSACWQDGRFAENETWVYDSCCTCTCKKFKTICHQITC 355
DB 300 KYTEENKELANELRRP-----LCYHNGVQYRNNEWTVDSCTECHQNSVTICKVSC 353
QY 356 PPATCASPFVEGECPCSLSHSYDGEWSPWAEWTQCSVTGSGTQQRGRSCDVTSNYC 415
DB 354 PIMPENATVPDCECCPRCWPSDSADGWSWSEWTSCSTSCGNGIQQRGRSCDSLNNRC 413
QY 416 LGPSIQTRACLSKCDTRIDQGGWHSWSPWSSCSTCGVGNITRILCNLSPVPMGGKN 475
DB 414 EGSVQTRTCHIQECKRFKQDGGWHSWSPWSSCSTCGVGNITRILCNLSPVPMGGKN 473
QY 476 CKSGGRTACQACAPCIDGRWSPWSPWACTVTCAGGIRTRVRCNSPEPQYGGKACV 535
DB 474 CEGARETRACKDACPINGWGPWSPWDICSVTCGGVQKRSRLCNNTPTPQGGKDCVG 533
```

```
QY 536 DYQEROMCNKRSQPYDGLCLSNPCFPGAQCSSRPPDGSSWSCGCPVGLNGHTHCELDCA 595
DB 534 DVTENOICNKQDCPIDGCLSNPCFAGVKTSTPDGSKWCKACPPGSGNGIQCTDVECK 593
QY 596 LVPDICFSTSKVPRCVNTQGFHCLPCPPRYRGNQVGVGLEAAKTEKQVCEPENCKDK 655
DB 594 EYVDAFCNHEHRCENTDPGYNCLPCPPRFTGSGQFPGQVEHATANKQVCKRNPCTDG 653
QY 656 THNCHKHAECIYLGHFSDPMYKCECQTGAGDGLICGEDSLDGLWNLVLCATNATYHC 715
DB 654 THDCNNKACNYLGHYSDPMYRCEKPGYAGNGIICGEDTDLGWPNEHLVVCANATYHC 713
QY 716 IKDNCPLHNSGOEDFDKDGIGDACDDDDNDGVTDEKNCQLLFPNPRADYDKDEVGDR 775
DB 714 KDCNCFNLNSQGEDYDKDGIGDACDDDDNDKIPDRONCPHYNPAQYDYDRDDVGD 773
QY 776 CNDPYYVHNPAQIDTNNEGEGDACSVDIDGDDVFNERNDCNPPYVNTDQDRDGDGVGDHC 835
DB 774 CNDPYYNHPDQADTNNEGEGDACAADIDGDLNLNERDNCQYVYVNDQDRDMDGVGDQC 833
QY 836 DNCPLVHNPDOQTDVNDLVGDCDDNNEDIDDGCHQNNQDNCNCPYISNANOADHRCOGDA 895
DB 834 DNCPLVHNPDOQTDVNDLVGDCDDNNEDIDDGCHQNNQDNCNCPYISNANOADHRCOGDA 893
QY 896 CDPDDNDGVPDDDRNCLRVFNPQDELDGDRGDKDDDFDNDNIPDIDDVCPENNAIS 955
DB 894 CDHDDNDGIPDDKONCLRVNPQDKSDGDRGDKDDDFDNDNIPDIDDVCPENNAIS 953
QY 956 EYDFRNFQWVPLDKTQIDPNWVIRHOGKELVQVANSDPGATGAFDEFGVDFSGTFY 1015
DB 954 EYDFRNFQWVPLDKTQIDPNWVIRHOGKELVQVANSDPGATGAFDEFGVDFSGTFY 1013
QY 1016 VNTDRDDYAGVFGYSSSRFVVMKQVOTYQWEDQTPRAYGSGVSLKVVNSTTGTG 1075
DB 1014 INTERDDYAGVFGYSSSRFVVMKQVOTYQWEDQTPRAYGSGVSLKVVNSTTGTG 1073
QY 1076 EHLRNALWHTGNTPGQVRLTWHDPNIGMKDYATYRHLTHRPKTYIRVLVHEGQVMA 1135
DB 1074 EHLRNALWHTGNTPGQVRLTWHDPNIGMKDYATYRHLTHRPKTYIRVLVHEGQVMA 1133
QY 1136 DSGPIYDQYAGRGLGLFVFSQEMVYFSDLYKCECD 1171
DB 1134 DSGPIYDQYAGRGLGLFVFSQEMVYFSDLYKCECD 1169
```

RESULT 6
TSPL_BOVIN
ID TSPL_BOVIN STANDARD; PRT; 1170 AA.
AC Q28178; Q28179;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Thrombospondin 1 precursor.
GN THBS1 OR TSPL OR TSP-1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HOLSTEIN; TISSUE-Tooth;
RX MEDLINE=98173773; PubMed=9507054;
RA Ueno A., Yamashita K., Nagata T., Tsurumi C., Miwa Y., Kitamura S.,
Inoue H.;
RT "cDNA cloning of bovine thrombospondin 1 and its expression in
RT odontoblasts and preodontoblasts";
RL Biochim. Biophys. Acta 1382:17-22(1998).
RN [2]
RP SEQUENCE OF 1-18 AND 710-1170 FROM N.A.
RC TISSUE-Aortic endothelium;
RA Zafar R.S., Moll Y.D., Womack J.F., Walz D.A.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.

-1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN, LAMININ, TYPE V COLLAGEN AND INTEGRINS ALPHA-V/BETA-1, ALPHA-V/BETA-3 AND ALPHA-IIb/BETA-3. MAY PLAY A ROLE IN DENTINOGENESIS AND/OR MAINTENANCE OF DENTIN AND DENTAL PULP.

-1- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.

-1- TISSUE SPECIFICITY: ODONTOBLASTS.

-1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.

-1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.

-1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.

-1- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AB005287; BAA21115.1; -
 EMBL; X87618; CAA60950.1; -
 EMBL; X87619; CAA60951.1; -
 HSP; P35555; IEMO.
 GlycoSuiteDB; Q28178; -
 InterPro; IPR000561; EGF-like.
 InterPro; IPR001881; EGF_Ca.
 InterPro; IPR001791; Laminin_G.
 InterPro; IPR000884; TSP1.
 InterPro; IPR003129; TSPN.
 InterPro; IPR001007; VWFC.
 InterPro; IPR003367; tsp_3.
 InterPro; IPR003367; tsp_3.
 Pfam; PF02210; TSPN; 1.
 Pfam; PF02412; tsp_1; 3.
 Pfam; PF00090; tsp_1; 3.
 Pfam; PF00093; wvc; 1.
 SMART; SM00181; EGF; 3.
 SMART; SM00209; TSP1; 3.
 SMART; SM00210; TSPN; 1.
 SMART; SM00214; WVC; 1.
 PROSITE; PS00022; EGF_1; FALSE_NEG.
 PROSITE; PS01186; EGF_2; FALSE_NEG.
 PROSITE; PS01208; VWFC; 1.
 PROSITE; PS50092; TSP1; 3.
 Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
 EGF-like domain; signal.
 SIGNAL 1 18
 CHAIN 19 1170
 DOMAIN 19 232
 DOMAIN 316 373
 DOMAIN 379 430
 DOMAIN 435 491
 DOMAIN 492 548
 DOMAIN 549 587
 DOMAIN 588 645
 DOMAIN 646 690
 DOMAIN 723 758
 DOMAIN 759 781
 DOMAIN 782 817
 DOMAIN 818 840
 DOMAIN 841 878
 DOMAIN 879 914
 DOMAIN 915 950
 DOMAIN 951 1170
 C-TERMINAL
 CELL ATTACHMENT SITE (POTENTIAL).
 INTERCHAIN (PROBABLE).
 INTERCHAIN (PROBABLE).
 BY SIMILARITY.
 BY SIMILARITY.
 BY SIMILARITY.
 BY SIMILARITY.
 BY SIMILARITY.
 BY SIMILARITY.

FT	DISULFID	620	644	BY SIMILARITY.
FT	DISULFID	650	663	BY SIMILARITY.
FT	DISULFID	657	676	BY SIMILARITY.
FT	DISULFID	678	689	BY SIMILARITY.
FT	CARBOHYD	248	248	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	360	360	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	708	708	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1067	1067	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1085	1085	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CONFLICT	805	805	S -> G (IN REF. 2).
SQ	SEQUENCE	1170 AA;	129533 MW;	ODD6ADF3E5FA031A CRC64;

Query Match 63.9%; Score 4223; DB 1; Length 1170;
 Best local Similarity 61.3%; Pred. No. 1.le-356;
 Matches 723; Conservative 172; Mismatches 267; Indels 14; Gaps 6;

QY	1	MVRL-VLLALWVWPSTQAGHQDKD-TTFDLFSISNTRKRTIGAKQFRGDPGPVAYRFV	58
DB	3	LANGLVLLLLHACGSRIPESGDNVDFIFELTGAAKRSGRLVKGDPSPAPRIE	62
QY	59	RFDYIPVNPADLSKITKIMROKEGFFLTALQKDGKSRGTLLALEGPGLSQROFEIVSN	118
DB	63	DANLIPVPDKKFDQLVDVAVRAEKGFLLASLRQMKTRGTLLAVERKDHSGVFSVSN	122
QY	119	GPADTLDTYWDGTRHVSLEVDGLADSKWKNVTVVAGETYSLVHVCCLDSFALDEP	178
DB	123	GKAGTLDLSLVGQKHWVSVEALLATGQWKSITLFEVDRAQLYIDCKEMNAELDVP	182
QY	179	---FYEHQLQAEKSMYVAKGSAHESHPGGLLVNHLVFNSEVDILSKGCGQGGAEIN	235
DB	183	IQSIFTRDLASIAKRLIAKGVND-NEQVLQNVRFVGTTPEDILNKGCSSTSVFT	241
QY	236	AISENTEITLGLPHVTTEYVGPSSERPVEVCESEELGNMVQELSLHLVNOPSENK	295
DB	242	L--DNNVVGSSPAIRTDYIGHKTKDLQAICGISCDELSSMLVLELRLRTVTTLQDSIR	299
QY	296	RVSNDNQFLWELIGGPPKTRNMSACWDGRRFAENETWYVDSTCTCTCKKFKTCHQITC	355
DB	300	KVTEENKELANELRRP-----LCYHNGVOYRTGDEWTVDSCTECRQNSVTICKKVC	353
QY	356	PPATCASPSFVEGECPCSLHSDVEGEGSWPWAETCCSVTCGSGTQQRGRSCDVTNCT	415
DB	354	PIMPCSNATVPDGECCPRCPWPSDSADGNSPSEWTSCTVTCGNGIQORGRSCDSLNR	413
QY	416	LGPSIQTRAGLSKCDTRIRQDGGWSHSPWSSCVTCGVGNITRIRLCHNSPVPQMGKN	475
DB	414	EGSSVQTRTCHIQCCKRRFKQDGGWSHSPWSSCVTCGSGVITRIRLCHNSPSPQMGKP	473
QY	476	CKSGRETAKACQAPCIDGRWSPWSPWACTVTCAGGIRERTRVCHSPSPQYGGKACVG	535
DB	474	CEGKARETKACQKDCPCINGWGPWSPWDICSVTCGGGVQKRSRLCHNKPQFGKDCVG	533
QY	536	DVEROMCNKRSPPVDGCLSNPCFPAGCSFSDGWSGSCGCPVGLNGTHCDBLDECA	595
DB	534	DVTENQICNKQDCPIDGCLSNPCFAGVQCTSYPDGSKKCCACPGYSGDVECKDVECK	593
QY	596	LVPDICEFSTSKVPRCVNTQPFHCLPCPPRYRGNQPVGVGLEAAKTEKQVCEPENCKDK	655
DB	594	EVDPACFNHNGHRCENTDPGYNCLPCPPRFTSQSPFGRGVHATANKQVCKPRNPCTDG	653
QY	656	THNCHKAECLYLGHFSDPMYKCEQGYAGDGLICGEDSDLDGWPNLNLVCATNATYHC	715
DB	654	THDCNNAKCNYLGHYSDPMYRCEKPGYAGNGLICGEDYDLGWPNEEDLVCVANATYHC	713
QY	716	IKONCPHLNPSGGDFDKDGGIGDACDDDDNDGVTDEKDNQCLLFNPRQADYDKDEVDGR	775
DB	714	RKNCNPLNPSGGEDYDKDGGIGDACDDDDNDKIPDDRDCNPFHYNPAQYDYDRDDVDGR	773
QY	776	CDNCPYVHNPAQIDTDNNGSGDACSVIDGDDVFNERNDCNPPYVNTQRTDGDGCVGDHC	835
DB	774	CDNCPYHNPDQADTDNNGEGDCAADIDGDSILNRDNCQYVYVNDQKDTMDMGVDGQC	833

QY 836 DNGPLVHPDQTDVNDLVGDCDNNEDIDDDGHQNNQDPCVYSNANOADHDDGQDA 895
 DB 834 DNGPLHNPDLQSDSDRIGDTCDDNDDIDDDGHQNNQDPCVYSNANOADHDDGQDA 893
 QY 896 CDPPDDNDGVPDRDRCLRVFNPDQEDLDGDRGDKDCKDDFNDNDIPDIDDDVCPENNALS 955
 DB 894 CDHDDNDGIPDDRDRCLRVFNPDQEDLDGDRGDKDCKDDFNDNDIPDIDDDVCPENNALS 953
 QY 956 ETDFRFQVPLDPKGTQTDIDPNWVTRHOGKELVQVANSDPGLAVGDFGSDVDSGTYY 1015
 DB 954 ETDFRFQVPLDPKGTQTDIDPNWVTRHOGKELVQVANSDPGLAVGDFGSDVDSGTYY 1013
 QY 1016 VNTDRDDVAGFVFGVQSSRFVVMKQVOTIYEDQTPRAYGSGVSKVYNSTGTG 1075
 DB 1014 INTERDDVAGFVFGVQSSRFVVMKQVOTIYEDQTPRAYGSGVSKVYNSTGTG 1073
 QY 1076 EHLRNALWHTGNTGGVRLTWDHDPNIGWMDYATYRHLTHRPKTYIRVLVHGGQVMA 1135
 DB 1074 EHLRNALWHTGNTGGVRLTWDHDPNIGWMDYATYRHLTHRPKTYIRVLVHGGQVMA 1133
 QY 1136 DSGPIYDQYAGRGLEFVFSQEMVFSDLKYECD 1171
 DB 1134 DSGPIYDQYAGRGLEFVFSQEMVFSDLKYECD 1169

RESULT 7
 TSPL_MOUSE
 ID TSPL_MOUSE STANDARD; PRT: 1170 AA.
 AC P35441;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Thrombospondin 1 precursor.
 GN THBS1 OR TSPL.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RP [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=92128941; PubMed=1774063;
 RA Lawler J., Duquette M., Ferro P., Copeland N.G., Gilbert D.J.,
 RA Jenkins N.A.;
 RT "Characterization of the murine thrombospondin gene.";
 RL Genomics 11:587-600(1991),
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92147683; PubMed=1371115;
 RA Laherty C.D.; O'Rourke K., Wolf F.W., Katz R., Seldin M.F.,
 RA Dixit V.M.;
 RT "Characterization of mouse thrombospondin 2 sequence and expression
 during cell growth and development.";
 RL J. Biol. Chem. 267:3274-3281(1992).
 RN [3]
 RP SEQUENCE OF 1-490 FROM N.A.
 RX MEDLINE=90375546; PubMed=2398070;
 RA Bornstein P., Alfi D., Devarayalu S., Framson P., Li P.;
 RT "Characterization of the mouse thrombospondin gene and evaluation of
 the role of the first intron in human gene expression.";
 RL J. Biol. Chem. 265:16691-16698(1990).
 CC -|- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
 CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
 LAMININ, TYPE V COLLAGEN AND INTEGRINS ALPHA-V/BETA-1, ALPHA-
 V/BETA-3 AND ALPHA-11B/BETA-3.
 CC -|- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.
 CC -|- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
 CC -|- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
 CC -|- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
 CC -|- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
 CC -|- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M62470; AAA50611.1; JOINED.
 DR EMBL; M62450; AAA50611.1; JOINED.
 DR EMBL; M62451; AAA50611.1; JOINED.
 DR EMBL; M62452; AAA50611.1; JOINED.
 DR EMBL; M62453; AAA50611.1; JOINED.
 DR EMBL; M62454; AAA50611.1; JOINED.
 DR EMBL; M62455; AAA50611.1; JOINED.
 DR EMBL; M62456; AAA50611.1; JOINED.
 DR EMBL; M62457; AAA50611.1; JOINED.
 DR EMBL; M62458; AAA50611.1; JOINED.
 DR EMBL; M62459; AAA50611.1; JOINED.
 DR EMBL; M62460; AAA50611.1; JOINED.
 DR EMBL; M62461; AAA50611.1; JOINED.
 DR EMBL; M62462; AAA50611.1; JOINED.
 DR EMBL; M62463; AAA50611.1; JOINED.
 DR EMBL; M62464; AAA50611.1; JOINED.
 DR EMBL; M62465; AAA50611.1; JOINED.
 DR EMBL; M62466; AAA50611.1; JOINED.
 DR EMBL; M62467; AAA50611.1; JOINED.
 DR EMBL; M62468; AAA50611.1; JOINED.
 DR EMBL; M62469; AAA50611.1; JOINED.
 DR EMBL; M87276; AAA53063.1; JOINED.
 DR EMBL; J05606; AAA40431.1; JOINED.
 DR EMBL; J05605; AAA40431.1; JOINED.
 DR PIR; A40558; A40558.
 DR PIR; B42587; B42587.
 DR PIR; A37905; A37905.
 DR HSSP; P35555; 1EMN.
 DR MGD; MGI:98737; Thbs1.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR001791; Laminin_G.
 DR InterPro; IPR000884; TSP1.
 DR InterPro; IPR003129; TSPN.
 DR InterPro; IPR001007; VWFC.
 DR InterPro; IPR003367; tsp_3.
 DR Pfam; PF00008; EGF; 2.
 DR Pfam; PF02210; TSPN; 1.
 DR Pfam; PF00090; tsp_1; 3.
 DR Pfam; PF02412; tsp_3; 8.
 DR Pfam; PF00093; VWC; 1.
 DR SMART; SM00181; EGF; 3.
 DR SMART; SM00209; TSP1; 3.
 DR SMART; SM00210; TSPN; 1.
 DR SMART; SM00214; VWC; 1.
 DR PROSITE; PS00022; EGF_1; FALSE_NEG.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS50092; TSP1; 3.
 DR PROSITE; PS01208; VWFC; 1.
 KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
 EGF-like domain; Signal.
 FT SIGNAL 1 18
 FT CHAIN 19 1170
 FT DOMAIN 19 232
 FT DOMAIN 316 373
 FT DOMAIN 379 430
 FT DOMAIN 435 491
 FT DOMAIN 492 548
 FT DOMAIN 549 587
 FT DOMAIN 588 645
 FT DOMAIN 646 690
 FT DOMAIN 723 758
 FT DOMAIN 759 781
 FT DOMAIN 782 817
 FT DOMAIN 818 840
 FT DOMAIN 841 878
 FT DOMAIN 879 914
 THROMBOSPONDIN 1.
 HEPARIN-BINDING (POTENTIAL).
 VWFC.
 TSP TYPE-1 1.
 TSP TYPE-1 2.
 TSP TYPE-1 3.
 EGF-LIKE 1.
 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
 EGF-LIKE 3.
 TSP TYPE-3 1.
 TSP TYPE-3 2.
 TSP TYPE-3 3.
 TSP TYPE-3 4.
 TSP TYPE-3 5.
 TSP TYPE-3 6.


```
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF Ca.
DR InterPro: IPR001791; Laminin G.
DR InterPro: IPR000884; TSP1.
DR InterPro: IPR003129; TSPN.
DR InterPro: IPR001007; WFEC.
DR InterPro: IPR003367; tsp_3.
DR Pfam: PF00008; EGF; 2.
DR Pfam: PF02210; TSPN; 1.
DR Pfam: PF00090; tsp_1; 3.
DR Pfam: PF02412; tsp_3; 8.
DR Pfam: PF00093; vwc; 1.
DR SMART: SM00181; EGF; 2.
DR SMART: SM00001; EGF-like; 1.
DR SMART: SM0209; TSP1; 3.
DR SMART: SM0210; TSPN; 1.
DR SMART: SM00214; vwc; 1.
DR PROSITE: PS00022; EGF_1; FALSE_NEG.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS00092; TSP1; 3.
DR PROSITE: PS01208; WFEC; 1.
DR Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
KW EGF-like domain; Signal.
FT SIGNAL 1 22
FT CHAIN 23 1173
FT DOMAIN 23 235
FT DOMAIN 319 376
FT DOMAIN 382 433
FT DOMAIN 438 494
FT DOMAIN 495 546
FT DOMAIN 550 590
FT DOMAIN 591 648
FT DOMAIN 649 693
FT DOMAIN 726 761
FT DOMAIN 762 784
FT DOMAIN 785 820
FT DOMAIN 821 843
FT DOMAIN 844 881
FT DOMAIN 882 917
FT DOMAIN 918 933
FT DOMAIN 934 1173
FT SITE 929 931
FT DISULFID 554 565
FT DISULFID 559 575
FT DISULFID 578 589
FT DISULFID 595 611
FT DISULFID 602 620
FT DISULFID 623 647
FT DISULFID 653 666
FT DISULFID 660 679
FT DISULFID 681 692
FT CARBOHYD 135 155
FT CARBOHYD 158 158
FT CARBOHYD 250 250
FT CARBOHYD 363 363
FT CARBOHYD 705 705
FT CARBOHYD 711 711
FT CARBOHYD 1070 1070
SQ SEQUENCE 1173 AA; 130019 MW; A9F036D6516C0F24 CRC64;

Query Match 63.4%; Score 4189; DB 1; Length 1173;
Best Local Similarity 61.3%; Pred. No. 1.4e-254; Indels 34; Gaps 9;
Matches 726; Conservative 161; Mismatches 264;

QY 5 LVLLALWVPSTQAGHQ-----DKDTDFLSISINRKT-----IGAKQFRGPDGPV 53
DB 4 IFLLMLVMPQT---HQAESGNDNSVDFLFLTGYNRKAGSKPQGLHLVKGPDPSP 60
QY 54 AYRFVFDYIPPNVADLSKTKIMQKGEFFLTAOLKQDGKSRGFTLLALEGFLGSLQRF 113
DB 61 AVRIEDADLIPLPEDKFQDLDAIRADRGFILLATLRQAKSRGALLSVERKDGGHIF 120

114 EIVSNGPADTLDLTWIDGTRHWVLEVDGLADSQWKNVTVOVAGETYSLHVGCCLIDSF 173
121 SLISNGRATLDLSLGERKQOVSVEDAVLATGWTNITLQVEDRAQLYVGCNMENA 180
174 ALDEP-----FYEHQAESRMVYAKGARESHRGLLQNVHLVFNSEVEDILSKGCGQ 229
181 ELDPVIOKIFTENL-ASTAHLRAKGVKVD-NFQGVQLQNVRFVFGTTLEAILRNKGLSM 238
230 QCAEI---NAISENTETLRLGPHVTVEYVGPSSRRPEVCERSCEBELGNMVELSLHYL 286
239 TNSVITLONPVNGSS-----PAIRTNVIGHKTKDLQAVCGFGSCDDLKFLFAEMKGLRTL 292
287 VNQPSLENLRVSDNQFLWELIGGPPKTRNMSACQDQGRFFAENETWYVDSCTCTCKRF 346
293 VTTLKDOVTKETEKNELIAIV----TTPGVCLHNGVLHKNRDEWTVDSCTECTCONS 347
347 KTIHQITCPCPAYCASPSFEVEGECPCSLHSDVBEGSWFAEWTCQSVTCSSGTOQRGR 406
348 ATICRKVSCPLMPCNTATIPDGECCPCWPSADSDWSPSDWTPCSTVTCGHGTOQRGR 407
407 SCDVTSNTCLGPSIOTRACLSKCDTRIQRDGGSHSWSPSSCSTVTCGVGNITRIRLNS 466
408 SCDSLNPNCEGSSVOTRSCQIQDCKRFQDGGSHSWSPSSCSTVTCGSGQITRIRLNS 467
467 PVPQMGKCKGSGRETKACQCAPCPIDGRWSPSPWSACTVTCAGGIRTRVCSNPEP 526
468 PVPQLNGRQCEGEGRENKPCQKDCPINGQWGPWSLWDCTVTCGGGMQKRELNCNPKP 527
527 QYGGKACVGVDOEROMCNKRSCPDVGCCLSNPCFPAQCSFPDGGSWSCSCPVGFLNGT 586
528 QYEGKDCIGEPDSDQCNKQDCPIDGCLSNPCFAGVKCTSFIDGSKWCSGCPGYRNGI 587
587 HCEDLDECALVPDICEFSTSKVPRCVNTOPGFHCLPCPRYRGVQVGVGLEAKTEKQVC 646
588 TCKDIDCKEVPDCACTFLNGVHRCENTEPGYNCLPCPRFTGTQPGKGIEEAKANKQVC 647
647 EPEPNCKDKTHCHKHAECIYLGHFSDPMYKCECTQYAGDGLICGEDSLDGLWPNLNV 706
648 KPRNFCADGTHDCHKNARCIYLGHSYSDPMFRCERCPGYAGNGIICGEDTLDGWPENLT 707
707 CATNATYHCIKCNPLHNSGDEDFKDGIGACDDDDNDGVDTEKCNQLLPNPROAD 766
708 CVDNATYHCLKCNCPNLNSGQEDYDKDGMGACDKDDNDGILLDRDRCQFVYNPAQYD 767
767 YDKDEVGDRCDNCPVYHNPQAJDTDNNGEGDACSVDIDGDDVFNERNDCNPPYVNTDORT 826
768 YDRDDVGDRCDNCPVYHNPQAJDTDNNGEGDACSVDIDGDDVFNERNDCNPPYVNTDORT 827
827 DGDGVGDHCDNCPVYHNPQAJDTDNNGEGDACSVDIDGDDVFNERNDCNPPYVNTDORT 886
828 DKDGVGDHCDNCPVYHNPQAJDTDNNGEGDACSVDIDGDDVFNERNDCNPPYVNTDORT 887
887 HPRDGGDADCDPDDNDGVPDDRDNCLVFNPDQEDLDGGRGDKDKDFDNDNIPDIDD 946
888 HDKDKGACDCHDDNDGVPDDRDNCLVFNPDQEDLDGGRGDKDKDFDNDNIPDIDD 947
947 VCPENNAISETDFRNFQVPLDPKGTQIDPNWVIRHQKELVQVANSDPGLAVGFDEFG 1006
948 VCPENVEISTDFRNFQVPLDPKGTQIDPNWVIRHQKELVQVANSDPGLAVGFDEFG 1007
1007 SYDFSCTGVNTDRDDYAGVFGYQSSSRFVVMKQVOTYTWEDQPTRAYGSGVSLK 1066
1008 AVDFSCTGVNTDRDDYAGVFGYQSSSRFVVMKQVOTYTWEDQPTRAYGSGVSLK 1067
1067 VYNSTGTGGEHLRNALWHTGNTPGQVRLTWHDPNRIGNKWDYATAYRHLHTRPKTGIRVL 1126
1068 VYNSTSGGGEHLRNALWHTGNTPGQVRLTWHDPNRIGNKWDYATAYRHLHTRPKTGIRVL 1127
1127 VHEGQVWADSGPIVDOTYAGGRLGLFVFSQEMVYFSDLKYECRD 1171
1128 MYEGKRVWADSGPIVDOTYAGGRLGLFVFSQEMVYFSDLKYECRD 1172
```

```

RESULT 9
COMP_HUMAN
ID COMP_HUMAN STANDARD; PRT; 757 AA.
AC P49747; Q16388; Q16389;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cartilage oligomeric matrix protein precursor (COMP).
GN COMP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cartilage;
RX MEDLINE=95229140; PubMed=7713493;
RA Newton G., Weremowicz S., Morton C.C., Copeland N.G.,
RA Gilbert D.J., Jenkins N.A., Lawler J.;
RT "Characterization of human and mouse cartilage oligomeric matrix
RT protein.";
RL Genomics 24:435-439(1994).
RN [2]
RP VARIANTS PSACH SER-459 DEL; TYR-468 AND TYR-472.
RX MEDLINE=95400301; PubMed=7670471;
RA Hecht J.T., Nelson L.D., Crowder E., Wang Y., Elder F.F.B.,
RA Harrison W.R., Francomano C.A., Prange C.K., Lennon G.G., Deere M.,
RA Lawler J.;
RT "Mutations in exon 17B of cartilage oligomeric matrix protein (COMP)
RT cause pseudoachondroplasia.";
RL Nat. Genet. 10:325-329(1995).
RN [3]
RP VARIANTS MED TYR-342, AND VARIANT PSACH ARG-328.
RX MEDLINE=95400302; PubMed=7670472;
RA Briggs M.D., Hoffman S.M.G., King L.M., Olsen A.S., Mohrenweiser H.,
RA Leroy J.G., Rimo D.L., Lachman R.S., Gaines E.S.,
RA Cekleniak J.A., Knowlton R.G., Cohn D.H.;
RT "Pseudoachondroplasia and multiple epiphyseal dysplasia due to
RT mutations in the cartilage oligomeric matrix protein gene.";
RL Nat. Genet. 10:330-336(1995).
RN [4]
RP VARIANT MED LYS-523.
RX MEDLINE=97173141; PubMed=9021009;
RA Ballo R., Briggs M.D., Cohn D.H., Knowlton R.G., Beighton P.H.,
RA Ramesar R.S.;
RT "Multiple epiphyseal dysplasia, ribbing type: a novel point mutation
RT in the COMP gene in a South African family.";
RL Am. J. Med. Genet. 68:396-400(1997).
RN [5]
RP VARIANT MED SER-371, AND VARIANT PSACH 513-VAL--LYS-516 DEL.
RX MEDLINE=97327574; PubMed=9184241;
RA Susic S., McGroarty J., Ahier J., Cole W.G.;
RT "Multiple epiphyseal dysplasia and pseudoachondroplasia due to novel
RT mutations in the calmodulin-like repeats of cartilage oligomeric
RT matrix protein.";
RL Clin. Genet. 51:219-224(1997).
RN [6]
RP VARIANTS PSACH AND MED.
RX MEDLINE=98130533; PubMed=9463320;
RA Briggs M.D., Mortier G.R., Cole W.G., King L.M., Golik S.S.,
RA Bonaventura J., Nuytink L., de Paape A., Leroy J.G., Biesecker L.,
RA Lipson M., Wilcox W.R., Lachman R.S., Rimo D.L., Knowlton R.G.,
RA Cohn D.H.;
RT "Diverse mutations in the gene for cartilage oligomeric matrix protein
RT in the pseudoachondroplasia-multiple epiphyseal dysplasia disease
RT spectrum.";
RL Am. J. Hum. Genet. 62:311-319(1998).
RN [7]
RP VARIANTS PSACH AND MED.
RX MEDLINE=99118868; PubMed=9921895;
RA Ikegawa S., Ohashi H., Nishimura G., Kim K.C., Sannohe A.,
RA Kimizuka M., Fukushima Y., Nagai T., Nakamura Y.;
RT "Novel and recurrent COMP (cartilage oligomeric matrix protein)
RT mutations in pseudoachondroplasia and multiple epiphyseal dysplasia.";
Hum. Genet. 103:633-638(1998).
RN [8]
RP VARIANTS PSACH AND MED.
RX MEDLINE=98112405; PubMed=9452026;
RA Loughlin J., Irven C., Mustafa Z., Briggs M.D., Carr A., Lynch S.-A.,
RA Knowlton R.G., Cohn D.H., Sykes B.;
RT "Identification of five novel mutations in cartilage oligomeric
RT matrix protein gene in pseudoachondroplasia and multiple epiphyseal
RT dysplasia.";
RL Hum. Mutat. Suppl. 1:S10-S17(1998).
RN [9]
RP VARIANT PSACH GLY-482.
RX MEDLINE=98112442; PubMed=9452063;
RA Susic S., Ahier J., Cole W.G.;
RT "Pseudoachondroplasia due to the substitution of the highly conserved
RT Asp482 by Gly in the seventh calmodulin-like repeat of cartilage
RT oligomeric matrix protein.";
RL Hum. Mutat. Suppl. 1:S125-S127(1998).
RN [10]
RP SUBUNIT: PENTAMER; DISULFIDE-LINKED.
CC (-) DISEASE: DEFECTS IN COMP ARE THE CAUSE OF PSEUDOACHONDROPLASIA
CC (PSACH) AND MULTIPLE EPIPHYSEAL DYSPLASIA (MED OR EDM), WHICH ARE
CC DOMINANTLY INHERITED CHONDRODYSPLASIAS CHARACTERIZED BY SHORT
CC STATURE AND EARLY-ONSET OSTEOARTHRITIS. MED IS BROADLY CATEGORIZED
CC INTO THE MORE SEVERE FAIRBANK AND THE Milder RIBBING TYPES. PSACH
CC IS MORE SEVERE AND IS RECOGNIZED IN EARLY CHILDHOOD.
CC (-) SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC (-) SIMILARITY: CONTAINS 4 EGF-LIKE DOMAINS.
CC (-) SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; L32137; AAA57253.1; -
CC EMBL; S79499; AAB35289.1; -
CC EMBL; S79500; AAB35270.1; -
CC HSSP; P02468; 1KLO.
CC MIM; 600310; -
CC MIM; 132400; -
CC MIM; 177170; -
CC InterPro: IPR000561; EGF-like.
CC InterPro: IPR001881; EGF_Ca.
CC InterPro: IPR003367; tsp_3.
CC Pfam; PF02412; tsp_3; 11.
CC SMART; SM00179; EGF_CA; 2.
CC SMART; SM00001; EGF_like; 2.
CC PROSITE; PS00022; EGF_1; FALSE_NEG.
CC PROSITE; PS01186; EGF_2; 1.
CC PROSITE; PS01187; EGF_CA; 2.
CC GlycoProtein; Cell adhesion; Calcium-binding; Repeat; EGF-like domain;
KW Signal; Disease mutation.
FT SIGNAL 1 20
FT CHAIN 21 757
FT POTENTIAL.
FT CARTILAGE OLIGOMERIC MATRIX PROTEIN.
FT N-TERMINAL.
FT DOMAIN 22 86
FT EGF-LIKE 1.
FT DOMAIN 87 126
FT EGF-LIKE 2.
FT DOMAIN 127 179
FT EGF-LIKE 3.
FT DOMAIN 180 222
FT EGF-LIKE 4.
FT DOMAIN 225 267
FT EGF-LIKE 5.
FT DOMAIN 297 332
FT TSP TYPE-3 1.
FT DOMAIN 333 355
FT TSP TYPE-3 2.
FT DOMAIN 356 391
FT TSP TYPE-3 3.
FT DOMAIN 392 414
FT TSP TYPE-3 4.
FT DOMAIN 415 452
FT TSP TYPE-3 5.
FT DOMAIN 453 488
FT TSP TYPE-3 6.
FT DOMAIN 489 524
FT TSP TYPE-3 7.
FT DOMAIN 525 757
FT C-TERMINAL.
FT DISULFID 69 69
FT INTERCHAIN (PROBABLE).
FT DISULFID 72 72
FT INTERCHAIN (PROBABLE).

```

```
FT DISULFID 91 102 BY SIMILARITY.
FT DISULFID 96 111 BY SIMILARITY.
FT DISULFID 114 125 BY SIMILARITY.
FT DISULFID 131 142 BY SIMILARITY.
FT DISULFID 136 151 BY SIMILARITY.
FT DISULFID 154 178 BY SIMILARITY.
FT DISULFID 184 197 BY SIMILARITY.
FT DISULFID 191 206 BY SIMILARITY.
FT DISULFID 209 221 BY SIMILARITY.
FT DISULFID 229 243 BY SIMILARITY.
FT DISULFID 237 253 BY SIMILARITY.
FT DISULFID 255 266 BY SIMILARITY.
FT CARBOHYD 121 121 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 742 742 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT SITE 367 369 CELL ATTACHMENT SITE (POTENTIAL).
FT VARIANT 290 290 D -> N (IN PSACH; MILD FORM).
FT VARIANT 299 299 /FTIQ-VAR_007614.
FT VARIANT 328 328 /FTIQ-VAR_007615.
FT VARIANT 342 342 C -> R (IN PSACH; MILD FORM).
FT VARIANT 349 349 D -> Y (IN MED; FAIRBANK TYPE).
FT VARIANT 361 361 D -> V (IN PSACH; MILD FORM).
FT VARIANT 361 361 D -> Y (IN MED).
FT VARIANT 367 368 /FTIQ-VAR_007620.
FT VARIANT 371 371 MISSING (IN MED).
FT VARIANT 372 372 /FTIQ-VAR_007621.
FT VARIANT 374 374 C -> S (IN MED; FAIRBANK TYPE).
FT VARIANT 387 387 /FTIQ-VAR_007622.
FT VARIANT 391 394 /FTIQ-VAR_007623.
FT VARIANT 408 408 /FTIQ-VAR_007624.
FT VARIANT 440 440 /FTIQ-VAR_007625.
FT VARIANT 440 440 /FTIQ-VAR_007626.
FT VARIANT 453 453 /FTIQ-VAR_007627.
FT VARIANT 459 459 G -> E (IN PSACH; MILD FORM).
FT VARIANT 468 468 /FTIQ-VAR_007628.
FT VARIANT 469 469 G -> R (IN PSACH; MILD FORM).
FT VARIANT 472 472 /FTIQ-VAR_007629.
FT VARIANT 472 472 N -> S (IN MED; FAIRBANK TYPE).
FT VARIANT 472 472 /FTIQ-VAR_007630.
FT VARIANT 472 472 MISSING (IN PSACH; SEVERE FORM).
FT VARIANT 472 472 C -> Y (IN PSACH; SEVERE FORM).
FT VARIANT 472 472 /FTIQ-VAR_007631.
FT VARIANT 472 472 /FTIQ-VAR_007632.
FT VARIANT 472 472 /FTIQ-VAR_007633.
FT VARIANT 472 472 D -> Y (IN PSACH; SEVERE FORM).
Query Match 30.2% Score 1993; DB 1; Length 757;
Best Local Similarity 54.1% Pred. No. 2.1e-117;
Matches 362; Conservative 73; Mismatches 176; Indels 58; Gaps 6;
QY 549 PVDGCLSNPCPGAQCSFFPDGWSNCGSPGVFLGNTHCEDLDCAVDPDICTFSTSKVP 608
Db 87 PLLHCAPGCFPGVACIQTESGG-RCGFCPAGFTGNGSHCTDVNCPNAHP--CEPR---V 140
QY 609 RCVNTQPGFHCPLPCPPRYRGNQPVGGLAAKTEKQVCEPENPCPKDKTHN----- 658
Db 141 RCINTSPFRCEACPGYSGTHQGVGLAFKANKQVCTDINECTGHOHNCVPSVCINT 200
QY 659 -----CHKHACVILYGHFSPMYKCEQT 682
Db 201 RGSFOGCPQPGVGDQASGQGRGAQRCPCDGSPECEHADCVL---ERDSRSCVCRV 257
```

```
QY 683 GYAGDGLIGEDSDLDGWPNNLVCATNATYHCIKDNCPLHNSGOEDFDKDGIGDACDD 742
Db 258 GWAGNGILCGRDYDLGFFDEKLRCEP---QCRKNCVTVPNSGQEDVDROGIGACDP 314
QY 743 DDNDGVTDEKNCQLLEFNPRQADYDKDEVGDRCDNCPYVHNPAQIDTDNNGEGDACSVD 802
Db 315 DADGDGVPEKNCPLVRNPQQRNTDEKWDGACDNCRSQKNDQDKDTQDQDGRGDACDD 374
QY 803 IDGDDVFNERNCPYVYNTDORTDGDGVDGCHDCNCPVHNPDOTDVDNDLVGDODDNE 862
Db 375 IDGDRIRNQADNCPVPNSDQSDGSDGIGDADNCPCQKSNPQADVDHDFVGDACDSQ 434
QY 863 DDDDDGHQNNQNCQPYISNANOADHRRDQDGDACDPPDDNDGVPDDRDNCRLVFNPDQED 922
Db 435 DODGDGHQSDRNCPCVPNSAQEDSDHDSQDGDACDODDDNDGVPDSRDNCRLVFNPDQED 494
QY 923 LGDGRGDIKDDDFDNDNIPDDDDVCPENNAISETDFRNFQMVPLPKGTQIDPNWVIR 982
Db 495 ADRDGVGVQDQDFADKVDKIDVCPENAEVTLTDFRAFTVVLDPEDGAQIDPNWVIL 554
QY 983 HOGKELVOTANSDDPGTAVGDFDEGSVDFTGVVNTDRDDYAGVFGVSSRFVYVNM 1042
Db 555 NQGREIVQTMNSDPGLAVGIFTAFNGVDFGTHTVNTVDDYAGFTFGYQDSSSFVYVNM 614
QY 1043 KQVQTQYWEQDPTRAYGSGVSLKVVNSTTGTGEHLRNALWHTGNTPPGOVRTLWHPDRI 1102
Db 615 KQMEQTYQANCFRAVAEPGIOLKAVKSTGEGEQLRNALWHTGDTESQVRLLRKDPNV 674
QY 1103 GWKDTATYRWHLTHRPKGTQYIRVLVHVGQVMDSGPIYDQYAGRLGLFVFSQBMVYF 1162
Db 675 GWKDKSKYRWFQLHRPQVGYIRVRYEGPELVADSNVLDTTMRGRLGVFCFSQBNIIW 734
QY 1163 SLDKYECRD 1171
Db 735 ANLRYRCND 743
RESULT 10
COMP_RAT STANDARD; PRT; 755 AA.
AC P35444.
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cartilage oligomeric matrix protein precursor (COMP).
GN COMP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cartilage;
RX MEDLINE=93054522; PubMed=1429587;
RA Oldberg A., Antonsson P., Lindblom K., Heinegaard D.;
RT "COMP (cartilage oligomeric matrix protein) is structurally related
to the thrombospondins."
RL J. Biol. Chem. 267:22346-22350(1992).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.05 ANGSTROMS) OF 27-72.
RX MEDLINE=97020111; PubMed=8864111;
RA Malashkevich V.N., Kammerer R.A., Efimov V.P., Schulthess T.,
RA Engel J.;
RT "The crystal structure of a five-stranded coiled coil in COMP: a
prototype ion channel?";
RL Science 274:761-765(1996).
CC -1- SUBUNIT: PENTAMER; DISULFIDE-LINKED.
CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC -1- SIMILARITY: CONTAINS 4 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL: X72914; CAA51419.1; -
 DR FIR; A44315; A44315.
 DR PDB; 1VDF; 08-OCT-97.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001881; EGF_Ca.
 DR InterPro: IPR003367; tsp_3.
 DR Pfam: PF02412; tsp_3; 9.
 DR SMART; SM00179; EGF_Ca; 2.
 DR SMART; SM00001; EGF-like; 2.
 DR PROSITE; PS00022; EGF_1; FALSE_NEG.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS01187; EGF_Ca; 2.
 KW Glycoprotein; Cell adhesion; Calcium-binding; Repeat; EGF-like domain;
 KW Signal; 3D-structure.
 FT SIGNAL 1 19
 FT CHAIN 20 755
 FT DOMAIN 21 84
 FT DOMAIN 85 124
 FT DOMAIN 125 177
 FT DOMAIN 178 220
 FT DOMAIN 223 265
 FT DOMAIN 295 330
 FT DOMAIN 331 353
 FT DOMAIN 354 389
 FT DOMAIN 390 412
 FT DOMAIN 413 450
 FT DOMAIN 451 486
 FT DOMAIN 487 522
 FT DOMAIN 523 755
 FT DISULFID 68 68
 FT DISULFID 71 71
 FT DISULFID 89 100
 FT DISULFID 94 109
 FT DISULFID 112 123
 FT DISULFID 129 140
 FT DISULFID 134 149
 FT DISULFID 152 176
 FT DISULFID 182 195
 FT DISULFID 189 204
 FT DISULFID 207 219
 FT DISULFID 227 241
 FT DISULFID 235 251
 FT DISULFID 253 264
 FT CARBOHYD 119 119
 FT CARBOHYD 740 740
 SQ SEQUENCE 755 AA; 52663 MW; AB48888FE093C598 CRC64;
 Query Match 29.8%; Score 1968; DB 1; Length 755;
 Best Local Similarity 54.4%; Pred. No. 7.6e-116;
 Matches 362; Conservative 73; Mismatches 179; Indels 52; Gaps 7;
 QY 549 PVDGCLSNPCFPGAGC-SSFPDGSWCGSCPVGFLGTHGTHCEDLDECAVPDIDCFSTSKVP 608
 DB 85 PVALCAPGSCFPGVCTETATGA-RCGCPGGYTGNGSHCTDVNECAHP--CFPR--V 138
 QY 609 RCVNTQPFHCLPCPRYRGNPQVGVGLENAKTEKQVCEPENCKDKTHNCHKABCI-- 666
 DB 139 RCINTSPGFCEACPFPGSGPTHEGVGLTFAKTNKQVCTDINECEGTQHCNCPVNSVCVNT 198
 QY 667 -----YLG-----HF-----SDPMKCECOTGVA 695
 DB 199 RGSFQCGPCQPGVGVQDRCQRRGHQFCPDGSPSPCKEAKADILERDGRSCVCAVGWA 258
 QY 686 GDGLIGEDSDLDGWFNLNVLVCATNATYHCIDKNCPLHNSGQEDFDKDGACDDDD 745
 DB 259 GNLGLCGRTDLGDFDEKLCRSE---QCRKNCVTVFNSGQEDVDRDRIGDCACPDAD 315

QY 746 NGVVTDEKNCQLLFPNRQADYDKDEVGRCDNCNCPYVHNPAQIDTDNNGEGDACSVDIDG 805
 DB 316 GDGVNPEQNCPLVRNPQDRNSDKWGDACDNCRCQKDDQKTDTRDGGQDADDDIDG 375
 QY 806 DDVFNERNDCPYVYNTDORDTDGVDGCHDCNCPVHNPDQTDVNDLVGDCQDNNEDID 865
 DB 376 DRIKRVADNCPVPVNFQSDSDGVDGVDACDNCPCQKNDPQDRVDHDFVGADCDSDQDD 435
 QY 866 DGHQNNQDNCPIYISNANQADHRCQGGDADCDPDDNDGVDPPDRDNCRLVFNPDQEDLDG 925
 DB 436 GDGHQDSRNCPTVPNSAQDSDHDKGKADCDDDNDGVDPSRDCRLVNPQEQDNDR 495
 QY 926 DGRGDIKDDFDNDNTPIDDDVCPENNAISETDFRNFQVPLDPKGTQTDPNWVIRHQG 985
 DB 496 DGVGDACQGDADKVIDKIDVCPENAEVTLTDFRAFTQTVLDPDEGDAQIDPNWVNLNOG 555
 QY 986 KELVQTANSDPGTAVGDFEGSVDFSGTFVYNTDRDDYAGVFGVQSSSRFYVMMKQV 1045
 DB 556 MEIVQWNSDPGLAVGTAFTAFGVDFEGTFHVNTATDDYAGVFGVQSSSRFYVMMKQV 615
 QY 1046 TOTYWEQDTPRAYGYSVSLKVNSTGTGTEHLRNALWHTGNTPGQVRTLWHDPRNTGWK 1105
 DB 616 EQTYQANPFAVAEPGCIQLKAVKSTGPGCEQLRNALWHTGDTASQVRLWKLKDPNRYGWK 675
 QY 1106 DYTAYRWHLTHRPKTYIRVVLVHEGQVWADSGPIVDQTYAGGRLGLFVFSQEMVYFSDL 1165
 DB 676 DKTSYRMFLQHRPQVGIYRIFYEPGLVADSNVLDATMRGKRLGVCFCSQENIIWANL 735
 QY 1166 KYECD 1171
 DB 736 RYRCND 741
 RESULT 11
 TSP3_HUMAN STANDARD; PRT; 956 AA.
 ID TSP3_HUMAN AC P49746;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Thrombospondin 3 precursor.
 GN THBS3 OR TSP3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 ON NCBI_Taxid=9606;
 RX NCBI_Taxid=9606;
 RT "Structure and organization of the human thrombospondin 3 gene
 (THBS3).";
 RL Genomics 27:329-336(1995).
 RN [2]
 RP SEQUENCE OF 1-736 FROM N.A.
 RX MEDLINE=97474796; PubMed=9331372;
 RA Winfield S.L., Tayebi N., Martin B.M., Ginns E.I., Sidransky E.;
 RT "Identification of three additional genes contiguous to the
 RT glucocerebrosidase locus on chromosome 1q21: implications for Gaucher
 RT disease";
 RL Genome Res. 7:1020-1026(1997).
 CC -!- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
 CC CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
 CC LAMININ AND TYPE V COLLAGEN.
 CC -!- SUBUNIT: OLIGOMER; DISULFIDE-LINKED.
 CC -!- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
 CC -!- SIMILARITY: CONTAINS 4 EGF-LIKE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

RT "Thrombospondin 3 (Tbbs3); a new member of the thrombospondin gene
 RL family.", Chem. 267:12192-12196(1992).
 RN [4]
 RP SEQUENCE OF 834-956 FROM N.A.
 RC STRAIN-BALB/C;
 RA Vos H.L., Mockingsturm-wilson M., Rood P.M.L., Maas A.C.E.,
 RA Duhig T., Gendler S.J., Bornstein P.;
 RA Submitted (Oct-1994) to the EMBL/GenBank/DBJ databases.
 RL -1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
 CC CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
 CC LAMININ AND TYPE V COLLAGEN.
 CC -1- SUBUNIT: OLIGOMER; DISULFIDE-LINKED.
 CC -1- TISSUE SPECIFICITY: BRAIN, LUNGS AND CARTILAGE.
 CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 4 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; L04302; AAA040497.1; -
 DR EMBL; L24434; AAA040433.1; -
 DR EMBL; M86620; AAA040430.1; JOINED.
 DR EMBL; M86612; AAA040430.1; JOINED.
 DR EMBL; M86613; AAA040430.1; JOINED.
 DR EMBL; M86614; AAA040430.1; JOINED.
 DR EMBL; M86615; AAA040430.1; JOINED.
 DR EMBL; M86616; AAA040430.1; JOINED.
 DR EMBL; M86617; AAA040430.1; JOINED.
 DR EMBL; M86618; AAA040430.1; JOINED.
 DR EMBL; M86619; AAA040430.1; JOINED.
 DR EMBL; U16175; AAA98537.1; -
 DR PIR; A46016; A46016.
 DR PIR; A44124; A44124.
 DR HSSP; P35555; 1EMN.
 DR MGD; MGI:98739; Tbbs3.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR003129; TSPN.
 DR InterPro; IPR003367; tsp.3.
 DR Pfam; PF02210; TSPN; 1.
 DR Pfam; PF02412; tsp.3; 8.
 DR SMART; SM00179; EGF_Ca; 2.
 DR SMART; SM00001; EGF_like; 2.
 DR SMART; SM00210; TSPN; 1.
 DR PROSITE; PS00022; EGF_1; FALSE_NEG.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS01187; EGF_Ca; 2.
 KW Glycoprotein; Cell adhesion; Calcium-binding; Repeat;
 KW EGF-like domain; Signal.
 FT SIGNAL 1 21
 FT CHAIN 22 956
 FT DOMAIN 22 272
 FT DOMAIN 274 315
 FT DOMAIN 316 369
 FT DOMAIN 370 413
 FT DOMAIN 414 456
 FT DOMAIN 488 523
 FT DOMAIN 524 546
 FT DOMAIN 547 582
 FT DOMAIN 583 605
 FT DOMAIN 606 643
 FT DOMAIN 644 683
 FT DOMAIN 684 719
 FT DOMAIN 720 956
 FT DISULFID 266 266
 FT INTERCHAIN (PROBABLE).

FT	DISULFID	269	269	INTERCHAIN (PROBABLE).
FT	DISULFID	278	289	BY SIMILARITY.
FT	DISULFID	283	300	BY SIMILARITY.
FT	DISULFID	303	314	BY SIMILARITY.
FT	DISULFID	320	332	BY SIMILARITY.
FT	DISULFID	326	341	BY SIMILARITY.
FT	DISULFID	344	368	BY SIMILARITY.
FT	DISULFID	374	388	BY SIMILARITY.
FT	DISULFID	382	397	BY SIMILARITY.
FT	DISULFID	400	412	BY SIMILARITY.
FT	DISULFID	418	432	BY SIMILARITY.
FT	DISULFID	426	442	BY SIMILARITY.
FT	DISULFID	444	455	BY SIMILARITY.
FT	CARBOHYD	310	310	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	407	407	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	644	644	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	937	937	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CONFLICT	433	433	L -> F (IN REF. 2).
FT	CONFLICT	563	563	N -> D (IN REF. 2).
FT	CONFLICT	720	720	G -> E (IN REF. 2).
FT	CONFLICT	871	871	L -> W (IN REF. 2 AND 4).
SQ	SEQUENCE	956 AA;	103972 MW;	BFFA666488C069A2 CRC64;

Query Match 29.3%; Score 1936.5; DB 1; Length 956;
 Best Local Similarity 52.0%; Pred. No. 9.4e-114;
 Matches 359; Conservative 81; Mismatches 181; Indels 69; Gaps 11;

QY	537	VBERQMC---	NKRCSPVDGCLSNPCFPQAQCS---	SFPDGSWSGSCPVGFLNGTHCED	590
DB	263	IMECQVCGHEQRS	----	HCSFSPCERGVDCMEVEYEP--	GYRCGCPGLQNGTHCDD
QY	591	LOECALVPDICTFSTKVPKVCNTQPGFHLCPKPPRYGNGQVPGVLEAAKTEKQVCE---	647		
DB	317	INECAHA-DPCFPSS---	CINTMPGFHCEACPPYKGTGTVSGVIGDIYARASQVNCID	372	
QY	648	-----	-----	PENCKDKTHN-CHKHAE	665
DB	373	ECNDGNGGCDPNISICTNTVSGFKGCRGLGFLGNSQGCVPARTCHSPAHSCHHAHC	432		
QY	666	IVLGHSFDPYKCECQTYAGDGLIGGEDSLDGWPNLNLVCATNATYHCKNCPLPN	725		
DB	433	LP---	ERNGAVSCOCNVHAGNVCVGPDTIDGYPDQALPCMDN-NKHCQDNCLLTPN	488	
QY	726	SGOEDFKDGIADACDDDDNDGVTDEKNCQLLFPNPRADYDKDEVGRDCNCPVHNP	785		
DB	489	SGOEDADNDGVGQCDADGDKINVEDNCRFLPNKQDQNSDTSFGDACDCNCPVNN	548		
QY	786	AQIDTDNNGEGDACSVDIDGDDVFNERNDCPVYNTDQRTDGDGVDGDCNCPVHNP	845		
DB	549	DQKDTDNGEGDACCNDNDVDGDIPIGLDNCVKVPNPLQTRDDEGVDGDCDCPEMSNPT	608		
QY	846	QTDVNDLVGDQCDNNEDIDDDGHQNNQNCVYISNANQADHRDQGGDACDPPDDNDGV	905		
DB	609	QTDADSLVGDVCDTNEGSDGSHQDTKNCPLPNSSQLSDNDGLGDECDGDDNDGV	668		
QY	906	PD---	DRNCRCLVFPNDQEDLDGDRGDIKDDDFDNDINPIDDDVCPENNAISDTRFN	961	
DB	669	PDYIPGPNCRCLVFPNPKQSDGNGVGDVCEDDFNDAVDVLDYCPESAGVTLDFRA	728		
QY	962	FQWVPLDPKTTQIDPNWVIRHOGKELVQTNDSDCPIANGVDFSGVDFSGTGYVNTDRD	1021		
DB	729	QYTVILDPGDAQIDPNWVNLQGMIEIVQTNDSPLAVGYTAFNGVDFEGTFHVTVD	788		
QY	1022	DDYAGVFYGYSSREYVYVWVKQVQTYWEDQPTRAYGYSGLSKVYVNSTGTGHLRNA	1081		
DB	789	DDYAGFLFSYQDSGREYVYVWVKQTEQTYWQATPFRAVAPGLKAVTSISGSGEHLRNA	848		
QY	1082	LWHTGNTPGQVRLWHDPRNIGWKDYATRWHLTHRPKGTGYRLVHESQWADSGPIY	1141		
DB	849	LWHTGHTPDQVRLWLDPRNVGLRDKTSYRWLLHRRPQVGYIRVKLYEGPQLVADSGVYI	908		
QY	1142	DQTYAGRLGLFVFSQEMVYFSDLKVECD	1171		

Db 929 IWSNLKYRCND 939

RESULT 14

TSP4_HUMAN STANDARD; PRT; 961 AA.

AC P35443;

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Thrombospondin 4 precursor.

GN THBS4 OR TSP4.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Heart;

RX MEDLINE=9335322; PubMed=9350346;

RA Lawler J., Duquette M., Uffry L., McHenry K., Smith T.F.;

RT "The evolution of the thrombospondin gene family.";

RL J. Mol. Evol. 36:509-516(1993).

CC -1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN, LAMININ AND TYPE V COLLAGEN.

CC -1- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.

CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.

CC -1- SIMILARITY: CONTAINS 4 EGF-LIKE DOMAINS.

CC -1- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; Z19585; CAA79635.1; -.

DR PIR; S36069; S36069.

DR HSSP; P02468; 1KLO.

DR TM; 600715; -.

DR InterPro: IPR000561; EGF-like.

DR InterPro: IPR001881; EGF_Ca.

DR InterPro: IPR003129; TSPN.

DR InterPro: IPR003367; tsp_3.

DR Pfam; PF02210; TSPN; 1.

DR Pfam; PF02412; tsp_3; 9.

DR SMART; SM00179; EGF_CA; 2.

DR SMART; SM00001; EGF_Like; 2.

DR SMART; SM00210; TSPN; 1.

DR PROSITE; PS00022; EGF_1; FALSE_NEG.

DR PROSITE; PS01186; EGF_2; 1.

DR PROSITE; PS01187; EGF_CA; 2.

KW Glycoprotein; Cell adhesion; Calcium-binding; Repeat; EGF-like domain;

KW Signal.

FT CHAIN 1 21

FT SIGNAL 22 961

FT CHAIN 22 961

FT DOMAIN 22 285

FT DOMAIN 286 325

FT DOMAIN 326 378

FT DOMAIN 379 419

FT DOMAIN 420 462

FT DOMAIN 492 527

FT DOMAIN 528 550

FT DOMAIN 551 586

FT DOMAIN 587 609

FT DOMAIN 610 647

FT DOMAIN 648 687

FT DOMAIN 688 723

FT DOMAIN 724 961

FT SITE 562 564

CELL ATTACHMENT SITE (POTENTIAL).

FT DISULFID 258 258 INTERCHAIN (PROBABLE).

FT DISULFID 261 261 INTERCHAIN (PROBABLE).

FT DISULFID 290 301 BY SIMILARITY.

FT DISULFID 295 310 BY SIMILARITY.

FT DISULFID 313 324 BY SIMILARITY.

FT DISULFID 330 341 BY SIMILARITY.

FT DISULFID 335 350 BY SIMILARITY.

FT DISULFID 353 377 BY SIMILARITY.

FT DISULFID 383 394 BY SIMILARITY.

FT DISULFID 388 403 BY SIMILARITY.

FT DISULFID 406 418 BY SIMILARITY.

FT DISULFID 424 438 BY SIMILARITY.

FT DISULFID 432 448 BY SIMILARITY.

FT DISULFID 450 461 BY SIMILARITY.

FT CARBOHYD 612 612 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 941 941 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 961 AA; 105801 MW; 55A48AF481AEB9DD CRC64;

Query Match 28.6%; Score 1890; DB 1; Length 961;

Best Local Similarity 53.1%; Pred. No. 7.7e-111;

Matches 355; Conservative 71; Mismatches 185; Indels 58; Gaps 9;

QY 549 PVDGCLSNPCFPGAQCSFPDGSWSCGSCPVGLNGTHCEDLDECALVPDICEFSTSKVP 608

DB 286 PPRCDSNPFCRGVQCSTDSRDG-FQCGPCPEGYTGNGITCIDVDECKYHP--CYPG---V 339

QY 609 RCVNTQPGFHCPCPPRYRGNPQVGLAAKTEKQVC-----EPEN----- 650

DB 340 HCINLSFGFCDACPVGFTGMVQGVGISFAKSNKQVCTIDECRNGACVPNSICVNTLG 399

QY 651 -----PCK-----DKTHNCHKHAEIYLGHFSDPMYKCEQCTGYAG 686

DB 400 SYRCGPKCYGTGDIQIRGCKVERNCRNPELNPQSVNAQCI---EERQGDVTCVCGVGWAG 456

QY 687 DGLICEDSLDGLWPNLVCATNATYHCKDKNCPLHNSGOEDFKDGLGDACDDDDN 746

DB 457 DGTICGKVDIDSYDEELPCFSAR---NCKKDKCYVPSGQEDADRDGIGDACEADG 513

QY 747 DGYTDEKDNQQLFNPRQADYDDEVGDRCDNCPYVHNPAQIDTDNNEGDACSVDIDG 806

DB 514 DGIILNEQDNCVLHNNVDQRNSDKDIFGDACDNCLSVLNNDQKDTDGDGRGDACDDMDG 573

QY 807 DVNERDNCYVYNTDQRTDGDGVGDHDCNCPVHNPDQTDVDNDLVGDCDNNEDID 866

DB 574 GIKNILDNCPKFPNRQDKDGDGVGDACDSCPDVSNPNQSDVDNDLVGDCDNTQSDG 633

QY 867 DGHONNODNCPYISNANQADHRDGDGACDPPDDNDGVDP---DRDNCRLVFNPDQED 922

DB 634 DGHQDSTDNCPYVNSAQLDTRDKDGDGDECDDDDDNDGIFDLVPPGPDNCRVNPQAQED 693

QY 923 LDGDRGDIKDDFDNDNIPDIDDVCPENNAISETDFRNFQMYPLDPKGTQTIDPNMVR 982

DB 694 SNSDGVGDICESDFDQDVIDRIVDVCPEAEVLTDFRAYQTVGLDPEGDAQIDPNMVL 753

QY 983 HQKELVQRTANSPDGTAVGDEFSGVDFSGTFYVNTDRDDYAGVFGYSSSRFYVMW 1042

DB 754 NQGMIEVQTMNSDPLGAVGYTAENGVDYFEGTFVNTQTDYDAGIFGYDSSSFYVMW 813

QY 1043 KOVYQYWEQDQPRAYGYSGLKVVNSTGTGTEHLRNALWHTGNTPGQVRLTWHDPRI 1102

DB 814 KQEQYVQWQTPRAVAEPQIQKAVKSKTGPGEHLNLSLWHTGSDTQVRLWLKDSRV 873

QY 1103 GWKDYTAIRWHLRPRPKTYGIRVLVHEGKQVMADSPYDQTYAGGRGLGFVFSQEMVYF 1162

DB 874 GWKDKYSRWFLQHRPQVGYIRVRYEGLSELVADSGVTIDTMRGGRGLGFCFSQENI 933

QY 1163 SLDKYECD 1171

DB 934 SNLYKRCND 942

RESULT 15

FT	DISULFID	348	359		BY SIMILARITY.
FT	DISULFID	353	368		BY SIMILARITY.
FT	DISULFID	371	395		BY SIMILARITY.
FT	DISULFID	401	412		BY SIMILARITY.
FT	DISULFID	406	421		BY SIMILARITY.
FT	DISULFID	424	436		BY SIMILARITY.
FT	DISULFID	442	456		BY SIMILARITY.
FT	DISULFID	450	466		BY SIMILARITY.
FT	DISULFID	468	480		BY SIMILARITY.
FT	CARBOHYD	631	631		N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	960	960		N-LINKED (GLCNAC. .) (POTENTIAL).
SO	SEQUENCE	980 AA;	108213 MW;	056D41EB6E206FCF CRC64;	
Query Match					
Best Local Similarity		28.5%;	Score 1881;	DB 1;	Length 980;
Matches 440; Conservative 124;		Mismatches 341;	Indels 290;	Gaps	
Qy	6 VLLALWMPSTQAGHQDKDTTFDLFSINRNRTIGAKFRGPDGPVAYRVRFDYIPP 65	: : :	:		
Db	28 LLLHLVLPWRAGAQAATPOVFDLLP-SSSQRNLPAALQ-----P 66	:			
Qy	66 VNADDLSKITKIMQKEGFPLTAQLKODGKSRSRGTLLIALEGSLSQRFELVSNGPADTLD 125	:::::	:		
Db	67 VLTDp-----TLHELXIVISTFKLQSKSSATIFGLYSDDNSKYFEETVMGRLNKAI 117	:			
Qy	126 LTW- IDGTRHVLSLEDVGGLADSQWKV----TVOVAGETYSLHWGCCLIDSF-ALDEP 178	:::::	:		
Db	118 LRYLKDGKIHLVVFNLIQJADRRHRLLRLLNQLRGAGSVELYLDCQVDVSNVLNPRA 177	:			
Qy	179 F---YEHLQAEKSMYVAKGSARESHFRGLLQNHLVFNSEVEDILSKKGCOQGGAEN 235	:::::	:		
Db	178 FSGLTQPQAIELRTFKRP-----QDFLEELKLVVRLGSLEQVASLQDCFLLQQSEPLA 230	:			
Qy	236 AISENTETLRLGPHVTTEYVGPSSERPVECBESCELGNMWQELSGLHLVLNQPSLENK 295	:::::	:		
Db	231 A-----TGTDGFNR-----OFLQMTO-----LNNLLGEVNDCLR 260	:			
Qy	296 RVSDNOFLWELIGPPPKTNMSACWDQGRFFAENETWYVDSCCTCTCKKFKTICHOITC 355	:			
Db	261 QOVKETSF-----RN-----PQCG-----TIACQCAC 279	:			
Qy	356 PPATCASPSPEBECPCSLHSVDGBEGWSFWAEWTQCSVTCCSGTQQRGRSCDVTSNTC 415	:::::	:		
Db	280 GPLSFQSP-----PNTLVPI-----APPAPPTPRRCDSSPCFRGVRC----- 319	:			
Qy	416 LGPSIQIIRACLSKCDPRIRODGGWSHWPSSCSVTGVGNITRIRLCNLSPVPMGGKN 475	:::::	:		
Db	320 -----TDR-----DG-----PQCG----- 329	:			
Qy	476 CKSGRETKACQAGAPCIDGRWSPWSPWSACTVTCAGGIHERTRVCNSPPQYGGKACVG 535	:::::	:		
Db	330 -----PCP-DG-----YTNGITCS----- 343	:			
Qy	536 DVQERQMKRSCPVDGCLNSCPFGAACSSFDPGWSGCSGCPVGFGL----- 583	:::::	:		
Db	344 -----DVDCKYHPCHPGVRCNTNLAG-FRCDACPVGTGPWMQGVGINFAK 389	:			
Qy	584 -NGTHCEDLDECALVPDICFESTSKVPRCVNTQPCFHLCPPRYRGNQPVGVGLEAAKTE 642	:::::	:		
Db	390 TNKVQCTDVDECR--NGACVLNS--ICINTLSYRCGPKPCYGTGDQTRG----- 435	:			
Qy	643 KQVCEPNPCKDXTNH--CHKHABCIYLGHFSDPMYKCEQTGYAG-DGLICGEDSDLGW 700	:::::	:		
Db	436 ---CRTERSERNPEONPCSYHAQCI---EEBROGDVTCVGVGWAGRAYVCGKDKVDIDS 489	:			
Qy	701 PNLNLVCATNATHCIKDNCPHLNPSGOEDFDRKDGDACDDDDDDNDGVTERDKNDCOLLF 760	:::::	:		
Db	490 PDELPCSAR---NCKKDNCKYPVNSQEADADRDRGIDGACDEDADGGGILNEQDNCVLTH 546	:			
Qy	761 NPROADYDKDEVDRDNCNCPYVNPAQIDTPDNNBEGDACSVDIGDDVFNERNDCPYVTN 820	:::::	:		
Db	547 NVDOIRNTDKDIFGDACDNCRGVLNNQKDYTDGGGKGDACDDDDMDGDKIKNILDNCFRVPN 606	:			

```

QY 821 TQORTDGDGVGDHCDNCPLVHNPQTDVNDLVGDQCDNNEIDDDGHQNNQDNCPIYS 880
Db 607 RQQRDRGDGVGDACDSCPDVSNPNQSDVDNDLVGSDCDTNQDSGDGHQDSTDNCPVI 666
QY 881 NANQADHRRDGGDADCPDDDDNDGVDP-----DRDNCRLVFNPDQEDLDGDRGDIKDDF 936
Db 667 NSAQLDTDKDGIQDECDDDDDNDGMPDLFPPGPDNCRCLVPNPAQEDSNNDGVGDIQAD 726
QY 937 DNDNTPDIDDVCPENNAISETDFRNFOMVPLDPKGTQIDPNWVIRHOGKELVQTANSDP 996
Db 727 DQKVIDRIDVCPENAEITLTDFAITQTVVLDPEGDAQIDPNWVILNQMEIVQTMNSDP 786
QY 997 GIAVGDFEFGYDFSGTFYVNTDRDDYAGFVFGYQSSRFYVVMKQVTOITYWEDQPTR 1056
Db 787 GLAVGYTAFNGVDFEGTFHVNTQDDDYAGFIFGYQSSRFYVVMKQTEQTYWQATPER 846
QY 1057 AYGYSGVSLKVYNSTGTGCEHLRNALWHTGNTPGQVRLWHDPRNIGWKDYTAYRWHLTH 1116
Db 847 AVAEPGIQLKAVKSKTGPGCEHLRNSLWHTGNTSDQVRLWLDKDSRNWGWKDKVSYRWFLQH 906
QY 1117 RPKTGYIRVLVHEGQVMADSGPIYDOTVAGGRGLGFVFSQEMVYFSDLKYECRD 1171
Db 907 RPQVGYIRVRYFEGSELVADSGVTIDTMRGGRLGVFCFSQENIIWSNLKYRCND 961

```

Search completed: August 9, 2002, 10:06:34
 Job time: 294 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2002, 10:01:19 ; Search time 46.31 Seconds
(without alignments)
4378.105 Million cell updates/sec

Title: US-09-822-682-2
Perfect score: 6605
Sequence: 1 MWRLVILALVWPSTQAGH.....VFSGEMVYFSDLKYECRDI 1172

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPREMBL_19:.*
1: sp.archaea:.*
2: sp.bacteria:.*
3: sp.fungi:.*
4: sp.human:.*
5: sp.invertebrate:.*
6: sp.mammal:.*
7: sp.mhc:.*
8: sp.organelle:.*
9: sp.phage:.*
10: sp.plant:.*
11: sp.rodent:.*
12: sp.virus:.*
13: sp.vertibrate:.*
14: sp.unclassified:.*
15: sp.rvirus:.*
16: sp.bacteriaph:.*
17: sp.archaeap:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1988	30.1	817	4	O14592
2	1966	29.8	755	11	O9R0G6
3	1959	29.7	755	6	O9B880
4	1913.5	29.0	851	13	O42507
5	1902	28.8	863	11	O9QYS3
6	1901	28.8	963	11	O921T2
7	1821	27.6	1060	5	O968S4
8	1805	27.3	1024	5	O9VM97
9	962	14.6	287	13	O91002
10	833	12.6	181	13	P79786
11	639.5	9.7	376	4	O15667
12	580	8.8	128	6	O28195
13	505	7.6	5636	4	O96RW7
14	504	7.6	2673	4	O96SC3
15	442	6.7	585	6	O9TTS5
16	429	6.5	1444	5	Q17591

17	418.5	6.3	4123	4	O75851
18	417.5	6.3	1637	6	O9XSV8
19	410	6.2	2327	13	O9IBG7
20	405	6.1	1062	5	O19204
21	400	6.1	1202	4	O9P283
22	399	6.0	788	5	O22631
23	330	5.9	89	11	O9R150
24	331	5.8	229	6	O28194
25	379	5.7	1121	5	O966P9
26	379	5.7	1148	5	O966Q0
27	378.5	5.7	1081	5	O9U631
28	377	5.7	103	6	O9N2C3
29	376.5	5.7	1083	5	O9VTT0
30	365.5	5.5	815	4	O96JS2
31	358.5	5.4	425	6	O02661
32	341	5.2	2189	5	O9BI05
33	341	5.2	2524	5	O9GPA5
34	333.5	5.0	807	6	O9GLX9
35	332.5	5.0	624	4	O94862
36	331.5	5.0	807	4	O9HCB6
37	322	4.9	802	13	O9W770
38	322	4.9	2471	11	O9QW30
39	321	4.9	2470	11	O35516
40	319.5	4.8	769	5	O00816
41	318.5	4.8	2531	5	O16004
42	317	4.8	1360	5	O9TYK4
43	312.5	4.7	861	11	O9QW58
44	312	4.7	2471	4	O04721
45	311.5	4.7	257	5	O966K7

ALIGNMENTS

RESULT 1
O14592 PRELIMINARY; PRT; 817 AA.
ID O14592
AC O14592;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE COMP_HUMAN.
GN COMP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Lamerdin J.E., McCready P.M., Adamson A.W., Burkhardt-Schultz K.,
RA Gordon L., Kyle A., Ramirez M., Stillwagen S., Garnes J., Panganan L.,
RA Christensen M., Bruce R., Quan G., Montgomery M., Ow D., Kobayashi A.,
RA Olsen A.O., Carrano A.V.;
RT "Sequence analysis of an -1 Mb region containing the MEF2B gene in
RT 19p12.";
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC003107; AAB86501.1; -;
DR HSSP; P35444; 1VDF.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR003367; tsp_3.
DR Pfam; PF02412; tsp_3; 11.
DR SMART; SM00179; EGF_CA; 2.
DR SMART; SM00001; EGF_like; 2.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 2.
KW Calcium-binding; EGF-like domain; Glycoprotein; Repeat.
SQ SEQUENCE 817 AA; 89148 MW; 70551980BE221855 CRC64;

Query Match 30.1%; Score 1988; DB 4; Length 817;

RESULT 3

Q9BG80 PRELIMINARY; PRT; 755 AA.

AC Q9BG80; 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE CARTILAGE OLIGOMERIC MATRIX PROTEIN.

OS COMP.

GN Equus caballus (Horse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.

OX NCBI_taxid=9796;

RN [1]

RP SEQUENCE FROM N.A.

RA Duthia J., Williams D., Smith R.K.;

RT "Molecular characterization and tissue distribution of equine

RT cartilage oligomeric protein (thrombospondin 5).";

RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF325902; AAG59881.1;

DR HSSP; P35444; 1VDF.

DR InterPro; IPR002048; EF-hand.

DR InterPro; IPR000561; EGF-like.

DR InterPro; IPR001881; EGF_CA.

DR InterPro; IPR003367; tsp_3.

DR Pfam; PF02412; tsp_3; 10.

DR SMART; SM00181; EGF_4.

DR SMART; SM00179; EGF_CA; 3.

DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.

DR PROSITE; PS01186; EGF_2; 1.

DR PROSITE; PS01187; EGF_CA; 2.

KW Calcium-binding; EGF-like domain; Matrix protein;

KW Repeat.

SQ SEQUENCE 755 AA; 82162 MW; 17CDFA4B1A79CE7 CRC64;

Query Match 29.7%; Score 1959; DB 6; Length 755;

Best Local Similarity 53.5%; Pred. No. 2.2e-146;

Matches 358; Conservative 73; Mismatches 180; Indels 58; Gaps 6;

QY 549 PVDGCLSNPCFPFGAQCSPFPDGSWCSGCPVGFGLNGTHCBEDLDECALVPDPCFSTKVP 608

DB 85 LAQCAPGSCFPVACVQTATGSA--RCGCPAGFTGNGPYCADVNECANP--CFPR---V 138

QY 609 RCVNTQPGFCLCPPPRYRGNQPVGVGLEAKTEKQCEPENCKDKTHN----- 658

DB 139 RCINTSPFRCCEACPPGTSGETHEGVGMFAKANKQVCTDIDECETGOHNCVPNSVCINT 198

QY 659 -----CHKHAECIYLGHFSDPMTKCECOT 682

DB 199 QGSFQCPCPGFVGDAQSGCRPRAQPCPDGTPSPCHEKADCYL---ERDGSRCVCVAV 255

QY 683 GYAGDGLICGEDSLDQWPNLVCATNAYHCITKDNCPHLPNSQGEDFDKDGIGDADCD 742

DB 256 GWAGNGLUCGRDRLDGFDPKELRCSE---QCRKDNCTVTPNSQGEDADRDGIGDADCT 312

QY 743 DDNDVGTDEKNCOLLFNPROADYDKDEVDGDCDNCPIYVHNPAQIDTDNNGEGDACSVD 802

DB 313 DADGCVPNEDGNCPLVNPDPQRTDQKKGADACDNCRSQKNDQKQDQDQGRGADACDD 372

QY 803 IDGDDVFNERNDCPIYVNTDQRTDQDGVGDHDCDNCPLVHNPQDQTDVNDLHVGQCDNNE 862

DB 373 IDGDIRNAVDNCPVNSDQKSDGIGDVCDCNCPQKSNPDQDQDHDHDFVGDACDSQ 432

QY 863 DIDDGHQNNQDNCPIYSNAQADHQRDGGDQDADPDDNDGVPDQDNCRLVNPQDQED 922

DB 433 DKDGDGHQSDRNCPIYNSAQDSDSGDQDADCDDEDNDNDGVPDSDRNCRLVNPQGED 492

QY 923 LDGDRGDRICKDFDNDNIPIIDDVCPENNAISFDFERNFQVPLDPKGTQTQIDPNVAVIR 982

DB 493 ADKRGVGDVCGQGFDAQKVDKIDVCPENNAEVLTLDFRAFQTVVLDPEGRQIDPNVAVL 552

QY 983 HQGKELVQTANSPGIAVGDFEFGSVDFSGTFYVNTDRDDDDYAGVFEGYQSSSRFFVVMW 1042

[illegible]

```
QY 788 IDTDNNGEGDASVDIDGDDVFNERNDCPYVYNTDQRTDGDGVDGHCNCPVHNPDQT 847
   III IIIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII
Db 446 KDTDANGEDACNDIDGDDGTPNMLDNCYPNPLQTDRLDGVGDACDSCPEASNPQA 505
   III IIIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII
QY 848 DVDNDLVGQDQNNEDIDDDGHQNNQDNCPIYSNANQADHRDGGDGDGDDNDGVDPD 907
   I:IIIIII I:IIII I:IIII I:IIII I:IIII I:IIII I:IIII I:IIII I:IIII I:IIII
Db 506 DADSLVAGDMCDTNDQRDGDGHQDKNCPIPNSSQLDSDNDGKGDGDCDDDDNDGIPD 565
   -----DRDNCRLVFNPDQEDLDGDDGDIKDDFDNDNDPIDDVCPENNAISETDFRNFQ 963
   IIIII I:IIII I:IIII I:IIII I:IIII I:IIII I:IIII I:IIII I:IIII I:IIII
Db 566 YIPGPDNCRILHNPQKSDSDGVDGVDCEEDFNDTVIDPLDVCPEAEVTLTDFRAYQ 625
   IIIII I:IIII I:IIII I:IIII I:IIII I:IIII I:IIII I:IIII I:IIII I:IIII
QY 964 MYPLDPKGTQIDPNMWIRHOGKELVOTANSDPGIAVDFGSDYDFSGTYFVNTDRDD 1023
   IIIII I:IIII I:IIII I:IIII I:IIII I:IIII I:IIII I:IIII I:IIII I:IIII
Db 626 TVILDPEGDAQIDPNWVNLNGMEIVQTMNSDPLGLAVGYTAFNGVDFEGTFHVNTVTD 685
   IIIII I:IIII I:IIII I:IIII I:IIII I:IIII I:IIII I:IIII I:IIII I:IIII
QY 1024 YAGFFVGQSSSRFVYVWVKQVTTYWEDQTPRAYGSGVSLKVYVNSTTGTGEHLRNALW 1083
   IIIII I:IIII I:IIII I:IIII I:IIII I:IIII I:IIII I:IIII I:IIII I:IIII
Db 686 YAGFFSYQDSSSFVYVWVKQVTTYWQATPFRAVAEPGLQLKAVKSTTGPEQLRNALW 745
   IIIII I:IIII I:IIII I:IIII I:IIII I:IIII I:IIII I:IIII I:IIII I:IIII
QY 1084 HTGNTPGQVRLHDPNRIGWKDYAYRWHLTHRPKTYIRVLVHEGQVMAVSGPIYDQ 1143
   IIIII I:IIII I:IIII I:IIII I:IIII I:IIII I:IIII I:IIII I:IIII I:IIII
Db 746 NYHTQDQVRLNWKDPRNVGHWKDTSYRWQLMHRPQVGYIRVLYEGVDLVADSGVLI 805
   IIIII I:IIII I:IIII I:IIII I:IIII I:IIII I:IIII I:IIII I:IIII I:IIII
QY 1144 TYAGRLGLFVSQEMYYFSLKYECD 1171
   IIIII I:IIII I:IIII I:IIII I:IIII I:IIII I:IIII I:IIII I:IIII I:IIII
Db 806 TMRGRLGVFCFSQENLIWSNLOYRCND 833
   IIIII I:IIII I:IIII I:IIII I:IIII I:IIII I:IIII I:IIII I:IIII I:IIII

RESULT 5
ID Q9QY53 PRELIMINARY; PRT; 863 AA.
AC Q9QY53;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
GN THROMBOSPONDIN 4 (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99431670; PubMed=10501972;
RA Newton G., Weremowicz S., Morton C.C., Jenkins N.A., Gilbert D.J.,
RA Copeland N.G., Lawler J.;
RT "The thrombospondin-4 gene.";
RL Mamm. Genome 10:1010-1016(1999).
DR EMBL; AF152393; AAD32714.1; -
DR EMBL; AF152392; AAD32714.1; JOINED.
DR HSSP; P35444; 1VDF.
DR MGD; MGI:110179; Thbs4.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR003129; TSPN.
DR InterPro; IPR003367; tsp_3.
DR Pfam; PF02412; tsp_3; 9.
DR SMART; SM00179; EGF_CA; 2.
DR SMART; SM00001; EGF_Like; 2.
DR SMART; SM00210; TSPN; 1.
DR PROSITE; PS00018; EF_HAND; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 2.
KW Calcium-binding; EGF-like domain; Glycoprotein; Repeat.
FT NON_TER
SQ SEQUENCE 863 AA; 95339 MW; 68E3EE6846728E4D CRC64;
```

Query Match 28.8% Score 1902; DB 11; Length 863;
Best Local Similarity 53.5% Pred. No. 8.7e-142;

```
Matches 358; Conservative 71; Mismatches 182; Indels 58; Gaps 10;
QY 549 PVDCLSNPCFPQAQCSFPGDSMSGSCPCVCFGLNGTHCEDLDCALVPDICEFSTKVP 608
   III IIIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII
Db 188 PTHCHDSSPCRGVRVCTDTRDG-FQCPCPDGYTGNGITCSVDVECKYHP--CYPG---V 241
   IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII
QY 609 RCVNTQPGFHLPCPPRYRGNQPGVGLAEAAKTEKQVC-----EPEN----- 650
   III IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII
Db 242 RCVNLAPGFRCDACPVGFTGPMVQGVGINFAKTNKQVCTDVEQCGNACVLNSICINTLG 301
   IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII
QY 651 -----PK-----DKTHNCHK-----HACIYLGHFSDPMYKCEQTGAG 686
   III III III III III III III III III III III III III III III
Db 302 SYRCGPCPKPGYTGDTGRCKTERSCRNPEQNPCSVHAQCI---EERQGDVTCVCGVGAG 358
   IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII
QY 687 DGLICGEDLDGPNLNLVLCATNATYHCIDKNCPHLNSGOEDFDKGIIGDADDDDDN 746
   IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII
Db 359 DGYVCGKDVDDISVPDEELPCSA---NCKDKNCKYVNSQGEDADRIGIGDADDEDAG 415
   IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII
QY 747 DGVTEKDNCKOLLNPNRQADYDKDEVDGDCNCPVYHNPQAIDTDNNGEGDACSVDIDGD 806
   IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII
Db 416 DGIILNEQDNCVLTHNIDQNSDKDIFGDACDNCRMVLNNDKQDGTGDCRGDADDDMDGD 475
   IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII
QY 807 DVFNERDNCPIYVNTDQRTDGDGVDGHCNCPVHNPDQTDVNDLVGDQCDNNEIDDD 866
   IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII
Db 476 GIKNILDNCPRVPHNDQDQDDVDGADCDSCPDVSNPNQSDVNDLVGDSCTNQDSDG 535
   IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII
QY 867 DGHQNNQDNCPIYSNANQADHRDGGQGDACDPDDNDNGVDP-----DRDNCRLVFNPDQED 922
   IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII
Db 536 DGHQDSTDNCPTVINSQDLDKDGIGDECDDDDDNDGIDPLVPPGPNCRCLVPNPAQED 595
   IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII
QY 923 LDGDRGDIKDDFDNDNIPDIDVCPENNAISETDFRNFQVPLDPKGTQIDPNWVIR 982
   IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII
Db 596 SNNDGVDGICEADFDQDQVIDHIDVCPENAEITLTDFRAYQTVVLDPEDAQIDPNWVVL 655
   IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII
QY 983 HOGKELVOTANSDPGIAVDFGSDYDFSGTYFVNTDRDDYAGVFCYQSSSRFVYVWV 1042
   IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII
Db 556 NQGMETVQTMNSDPLGLAVGYTAFNGVDFEGTFHVNTQDQDQDQDQDQDQDQDQDQ 715
   IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII
QY 1043 KQVTTYWEDQTPRAYGSGVSLKVYVNSTTGTGEHLRNALMHTGNTPGQVRLHDPNR 1102
   IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII
Db 716 KQTEQTYWQATPFRAVAEPGLKAVKSKTGPGEHLRNLNHTGTDTSQVRLWKKDSRV 775
   IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII
QY 1103 GWKDYATYRWHLTHRPKTYIRVLVHEGQVMAVSGPIYDQTYAGRLGLFVSQEMYYF 1162
   IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII
Db 776 GWKDKSVYRWFLQHRPQVGYIRVRYEGSELVADSGVTIDTMRGRLGVFCFSQENI 835
   IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII
QY 1163 SLDKYECD 1171
   IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII
Db 836 SNLKYRCND 844
   IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII

RESULT 6
ID Q921T2 PRELIMINARY; PRT; 963 AA.
AC Q921T2;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
GN THROMBOSPONDIN-4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99431670; PubMed=10501972;
RA Newton G., Weremowicz S., Morton C.C., Jenkins N.A., Gilbert D.J.,
RA Copeland N.G., Lawler J.;
RT "The thrombospondin-4 gene.";
RL Mamm. Genome 10:1010-1016(1999).
DR EMBL; AF102887; AAC73003.1; -
DR HSSP; P35444; 1VDF.
```


Db 531 VCDRAVCLMDNIRHKCHCNVWAGNGLICGRDITVDGWPQAIGC---PELHCQRDNC 587
 Qy 721 PHLPSGQEDFKGIGIGACDDDDNDGVDTEKNCQLLFPNQADYKDEVDGRDNC 780
 Db 588 PKLPSGQEDADLGGGDCDDADGNDVQNSQDNCWLAFTYQELDSDGDKYGVDCN 647
 Qy 781 YVNEPAQIDTNNNGEGDASVDIDGDDVFNERNDCNPPYVNTDQRTDGDGVDGHDNCPL 840
 Db 648 LKYNPQLDTEDEGLGDCDDIDNDSPNALDNCPLPNPSQSDVDNDGVDGDCDNC 707
 Qy 841 VHNPDQTDVNDLVDGQDNNEDIDDDGHQNNQDNCPIYSNANQADHDDGQDGDACDD 900
 Db 708 LNPQPKQDMDVFGDACHROIDGDDGVPSNLDNCPMVNSDQLDTDGDTGDCDDDM 767
 Qy 901 DNDGVDPRDNCRLVFNPDQEDLDGDRGDCIDKDFDNDNIPDIDVCPENNAISETDFR 960
 Db 768 DGDGIPNKNKCPALAKNPQDFNRNGKSCDEDDVDGVPMNDNCPNNSMIHHTDFR 827
 Qy 961 NFQMPVLPQKGTQIDPNVIRHOKELVQTANSDPGLAVGDFRSGVDFSGTFVYVNTDR 1020
 Db 828 TLQTIPLDPLKGLSQADPNVYVHANGTEIVQTLNSDPLGAVGKDAFGVDGTFYVNDT 887
 Qy 1021 DDYAGFVFGYSSRFYVVMKQVQTYWEDQPTRAYGSGVSLKVNSTTGTGHELRN 1080
 Db 888 DDYAGFVFGYSSRFYVVMKQVQTYWEDQPTRAYGSGVSLKVNSTTGTGHELRN 947
 Qy 1081 ALWHGNTPGQVRLTWHDPNRNIGWDTYAYRWHLRHRPKTYGIRVLVHVGQVMDSGPI 1140
 Db 948 SLWHEGNTDGEARLQPKNKPAWKERTSYRSLVHRPFAIGLIRLOMHEGNRLIFDSGNV 1007
 Qy 1141 YQQTAGGRLGFLVFSQEMVYFSLKYECRD 1171
 Db 1008 FDSLTKGRLGVCFPSQRMILWSNLOYKCN 1038
 RESULT 8
 Q9VM97 PRELIMINARY; PRT: 1024 AA.
 AC Q9VM97
 DT 01-MAY-2000 (TremBLrel. 13, Created)
 DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
 DE CG11326 PROTEIN.
 GN CG11326
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Chertis J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Paulos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieres S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Relneit K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 DR EMBL; AE003614; AAF52425.1; -;
 DR HSSP; P35555; 1EMN.
 DR FlyBase; FBgn0031850; Tsp.
 DR InterPro; IPR000152; Asx_Hydroxyl.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR001881; EGF-Ca.
 DR InterPro; IPR003367; tsp_3.
 DR Pfam; PF000008; EGF_3.
 DR Pfam; PF02412; tsp_3; 8.
 DR SMART; SM001179; EGF_CA; 1.
 DR SMART; SM00001; EGF-like; 3.
 DR PROSITE; PS001010; ASX_HYDROXYL; 1.
 DR PROSITE; PS01186; EGF_2; 2.
 DR PROSITE; PS01187; EGF_CA; 2.
 KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
 SQ SEQUENCE 1024 AA; 114398 MW; 88F279D6804610AD CRC64;

Query Match 27.3%; Score 1805; DB 5; Length 1024;

Best Local Similarity 34.08; Pred. No. 5, 4e-134;

Matches 407; Conservative 134; Mismatches 354; Indels 302; Gaps 31;

Qy 77 IMRQKGFLLTAQLKQDGKSGRTLLALEGPGLSQ-RQFEIVSNGPADLDTLYWDGTRH 135

Db 6 MLKHKMSFFL-----DRKQQRVTLDISANGATESRNFEPININETSTI----- 48

Qy 136 VVSLVDGLADSQWKNVTVOAGETSLHVGCDLLDSFALDEPFYEHQAESRMVAKG 195

Db 49 -----RSLAQFSKNRITLHVDCAKSTHHDID-----MNLAKLYTQMD 86

Qy 196 SARESHFRGLLQNVHLVFNSEVEDILSKKGQOQG---QGAELNAISENTEITLRGLPHV-- 250

Db 87 DPVIKLFRRKYPLH--FDGDMHSLQRANCOQKGNHRGNRMLNKITERGELFPRILG 144

Qy 251 ----TTEYV-----GPSSRRPEV---CER----- 268

Db 145 YLOPTMYIYSAWYAPIFAEKNKRDVRGWEPTIAREGVVDHRHQEVPTDVERGDIPLY 204

Qy 269 --SCEELNMVQELSLHVLVNPSENKRYNDNQFLWELIGGPKTRNMSACWDGRF 326

Db 205 NGDCEDA--LARSDDLALVKLLREDVAHQEITAYL-----RM 242

Qy 327 FAENETWTVDSCTTCT-CKKFKTICHQITC-PPATFASPSFVEGCC-----PSCLSHVD 379

Db 243 LLEN-----CAGCKNPLTNDQLRIEPCDRCANPCYPGVCELDASAAGPRCGH--- 289

Qy 380 GEEGSPWAEWTQCQSVTCGSGTQQRGRSCDVTNTCLGPSIQTRACSLSKDTRTRQDG 439

Db 290 -----CPLGFIGDGKSK-----PGV---TCAHMCYVPGVQ----- 317

Qy 440 WSHWSPWSSCSTVCVGVNITRILCNLSVPVPMGKNGKSGRETRACACGACPCIDRWSP 499

Db 318 -----CHDTV-----NGAQCD----- 328

Qy 500 WSPWSACTVTCAGGIRTRVCSNPSPQYGGKACVGDVQERQMCNKRSCPDGCLSNPCF 559


```
Db 329 -----SCPAGYEGDGRGTSLRNP-----CLDTPCP 353
QY 560 PGAQC---SSFPDGSWCGSPVGLNGTHCEDLDECALVPDICTFSTSKVPRCVNTQPGF 617
Db 354 SGAQCLQGVGPP-YFHCISCPMGHEVNGTSCRMNEC-LLYDPC---DELATCTNLSPGF 408
QY 618 HCLPCPRRYGNQPVGVGLE--AAKTEKQVCEPENPCDKTUNCHKHABCI-----666
Db 409 QCSPCPVGFDGTHAGFYADYSDVYRRQTCLDVDVDECRGFFRCPHEHSCINEIGSYRQ 468
QY 667 -YLGHSDDPMY-----KCECQTGYAGDGLICGD 694
Db 469 CHEGYVINGTYSCLDSSVFMCPDGTGVCDRNAVCLRMDNIRHKCHCNVAGNGLICGRD 528
QY 695 SLDJGWPNLNLVCATNATYHCKDKNCPLHNSQGEFDKDGIGDADCDDDDDNDGVTDK 754
Db 529 TDVGDWPDQAIGC---PELRCQORDNCPLKPLNSQGEADLDGHDGDCDDADGDNVQNSD 585
QY 755 NCQLLENPRQADYKDEVDGDRCDNCPCYVHNPAQIDTDNNGEGDACSVDIDGDDVFNERDN 814
Db 586 NCWLAYNTEQLSDSGDKVGDVCDNCVLYKNPRQLDTEDEGLGDECDGIDNDSIPNALDN 645
QY 815 CPVYVNTDQRTDGCVGDRCDNCPLVHNPDQTDVNDLVGQCDNNEIDDDGHONNOD 874
Db 646 CPLLPNSQSDVDNDGVGDACDNCPLPNPDQKDRDMDFVGACHRDIDGDDGVPNSLD 705
QY 875 NCPYISNANOADHRDQGDGACDPPDDNDGVPDRDNCRLVFNPDQEDLDGGRGDICKD 934
Db 706 NCPMVSNSQLDTDGDTGDECDDDMDGDIPIYKDCPLAKNPKQDDFNRRNGKGDSCD 765
QY 935 DFDNDNIPDIDVCPENNAISETDFRNFQVPLDPKGTQIDPNVIRQKGLVQTANS 994
Db 766 DEDVDGVPNGMDCNPNNSMIHHTFTLTQIDPLPKLSOADPNVYVHANGTEIVQTLNS 825
QY 995 DPGIAGVFDFEFGSVDFSGTFYVNTDRDDYAGVFYQSSRFYVYVWKKQVQTYWEDQP 1054
Db 826 DPLAVGKDAFGVDGTFYINDTDDYAGVFYQSSYKYVYVWKKGTQTYWEP 885
QY 1055 TRAYGSGVSLKVNSTTGTGEHLRNALWHTGNTQVTRLHDPNIGKWKDYATRWHL 1114
Db 886 FTASAPGIQIKLVNTEGPGPMRNSLWHEGNTDGEARLLWKDKPKNAWKERTSYRWSL 945
QY 1115 THRPKTYIRVLVKQYVADSGPIYDQTYAGRGGLGVFVSOEYVYVSDLYECRD 1171
Db 946 VHRPAIGLRLQWHEGNRLIFDSGNVDFSLKGLRGVFCFSORMIINSNLQIKCNN 1002
RESULT 9
ID Q91002 PRELIMINARY; PRT; 287 AA.
AC Q91002;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE THROMBOSPONDIN-4 (FRAGMENT).
GN Gallus gallus (Chicken).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauaria; Aves; Neognathae; Galliformes; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96016361; PubMed=7496039;
RA Tucker R.P., Adams J.C., Lawler J.;
RT "Thrombospondin-4 is expressed by early osteogenic tissues in the
chick embryo."
RL Dev. Dyn. 203:477-490(1995).
DR EMBL; L27263; AAA99960.1; -
DR InterPro; IPR003367; tsp_3.
DR Pfam; PF02412; tsp_3; 8.
FT NON_TER 1
```

```
FT NON_TER 287 287
SQ SEQUENCE 287 AA; 30880 MW; 9B3B0E413C95FC3F CRC64;

Query Match 14.6%; Score 962; DB 13; Length 287;
Best Local Similarity 61.6%; Pred. No. 5.1e-68;
Matches 175; Conservative 31; Mismatches 74; Indels 4; Gaps 1;

QY 761 NFRQADYKDEVDGDRCDNCPCYVHNPAQIDTDNNGEGDACSVDIDGDDVFNERDNCPYVYN 820
Db 4 MYNQNSDEDIFEGDADNCRSVLNNDQRTDGDGDKGACDDDDMDGDIKLLNLCORIPN 63
QY 821 TQORTDGVGDGDRCDNCPLVHNPDQTDVNDLVGQCDNNEIDDDGHONNODNCPYIS 880
Db 64 QQEDKNDNGVGDGDRCDNCPCYVSNPNOSDNDLVGSCDTPNQSDGHDGQDSTDNCPITII 123
QY 881 NANOADHRDQGDGACDPPDDNDGVPD----DRDNCRLVFNPDQEDLDGGRGDICKDDF 936
Db 124 NSSQLDTKDGLGDECDDEDDDGIPLDLPDPCDNCRLVFNPDQEDDNGDGVDCESDF 183
QY 937 DNDNIPDIDVCPENNAISETDFRNFQVPLDPKGTQIDPNVIRQKGLVQTANSOP 996
Db 184 DQDTVIDRIDVCPENAEITLDFRAYQTYVLDPEGAQIDPNVYVNLNQGMEIVQITNSDP 243
QY 997 GIAVGDFEFGSVDFSGTFYVNTDRDDYAGVFYQSSRFYV 1040
Db 244 GLAVGTAFNGVDGTFEGTFHVTVDYDAGFIFGYQDSSFYV 287

RESULT 10
P79786 PRELIMINARY; PRT; 181 AA.
AC P79786;
DT 01-MAY-1997 (Tremblrel. 03, Created)
DT 01-MAY-1997 (Tremblrel. 03, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE THROMBOSPONDIN-1 (FRAGMENT).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauaria; Aves; Neognathae; Galliformes; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97209337; PubMed=90556637;
RA Tucker R.P., Hagios C., Chiquet-Ehrismann R., Lawler J.;
RT "In situ localization of thrombospondin-1 and thrombospondin-3
transcripts in the avian embryo."
RL Dev. Dyn. 208:326-337(1997).
DR EMBL; U76994; AAB19208.1; -
DR InterPro; IPR003367; tsp_3.
DR Pfam; PF02412; tsp_3; 5.
FT NON_TER 1
SQ SEQUENCE 181 AA; 20202 MW; 365E8B0E97BEAE11 CRC64;

Query Match 12.6%; Score 833; DB 13; Length 181;
Best Local Similarity 75.6%; Pred. No. 4.2e-58;
Matches 136; Conservative 22; Mismatches 22; Indels 0; Gaps 0;

QY 686 GDGLICGEDSLDGPWNLNLVCATNATYHCKDKNCPLHNSQGEFDKDGIGDADCD 745
Db 1 GNGIICEEDTDLGWPENLVANATYHCKDKNCPLHNSQGEFDKDGIGDADCD 60
QY 746 NPGVTDEKDCOLLFPNPRQADYKDEVDGDRCDNCPCYVHNPAQIDTDNNGEGDACSVDIDG 805
Db 61 DDGIPDRDNCPCFYNPQQYDYDRDDVGDRCDCNCPYHNPNDEITDNNNGEGDACAVIDG 120
QY 806 DDVFNERNDCPYVYNTDQRTDGDGVDGDRCDNCPLVHNPDQTDVNDLVGQCDNNEID 865
Db 121 DGVLNERDNWQYVYVNDQRTDLDGVDGQCDNCPLEHNPNQDQEDTSDLDIGDECDNNQETD 180
```

```
RESULT 11
Q15667 ID Q15667 PRELIMINARY; PRT; 376 AA.
AC Q15667;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE THB51 PROTEIN PRECURSOR (FRAGMENT).
GN THB51.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86287276; PubMed=3461443;
RA Dixit V.M., Hennessy S.W., Grant G.A., Rotwein P., Frazier W.A.;
RT "Characterization of a cDNA encoding the heparin and collagen binding
RT domains of human thrombospondin.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:5449-5453(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92278426; PubMed=1950660;
RA Burglin T.R., Barnes T.M.;
RT "Introns in sequence tags [letter; comment].";
RL Nature 357:367-368(1992).
DR EMBL: M14326; AA61237.1;
DR InterPro: IPR001791; Laminin_G.
DR InterPro: IPR003129; TSPN.
DR InterPro: IPR001007; VWF.
DR Pfam: PF02210; TSPN; 1.
DR Pfam: PF00093; vwc; 1.
DR SMART: SM00210; TSPN; 1.
DR SMART: SM00214; vwc; 1.
DR PROSITE: PS01208; VWF; 1.
KW SIGNAL.
FT SIGNAL. 1 18
FT CHAIN. 19 >376
FT NON_TER 376
FT SEQUENCE 376 AA; 41329 MW; 313B62C0D253BA2B CRC64;

Query Match 9.7%; Score 639.5; DB 4; Length 376;
Best Local Similarity 34.4%; Pred. No. 2.7e-42;
Matches 137; Conservative 75; Mismatches 157; Indels 29; Gaps 7;

QY 1 MWRL-VLLALWVFWSTQGHQDKD-TTFDLFSININRKTIGAKQFRGPDGVPAYRFV 58
: ||| : | : : : ||| : : | : : ||| |||
Db 3 LAWGLVFLMHVCTNRIPESGDINSVDFIPELTGAARKGSRRLVKGPDPSAPFRIE 62
: ||| : | : : : ||| : : | : : ||| |||

QY 59 RFDYIPPNADLSKITKIMROKEGFFLTALQKDGKSRGTLTLLALEGPGLSQRFIYSN 118
: ||| : | : : : ||| : : | : : ||| |||
Db 63 DANLPPVPDDKFDQLVDVAERKGFLLASLRQMKKTGTLTALERKDHSGQVFSVSN 122
: ||| : | : : : ||| : : | : : ||| |||

QY 119 GPADTLDTLYIDGTRHRYVLESDVGLADSQMKNVTVQAGETYSYLVHGCDLIDSFALDEP 178
: ||| : | : : : ||| : : | : : ||| |||
Db 123 GRAGTLDSLTVQGHVVSVEALLATQWKSITLFQVEDRAQLYIDCKMENAEIDVP 182
: ||| : | : : : ||| : : | : : ||| |||

QY 179 F---YEHQAQKSRMYVAKGSRHFRGLLQNVHLVENSVEDILSKKGGQGGGASIN 235
: ||| : | : : : ||| : : | : : ||| |||
Db 183 IQSVETRLDASLARLIARLAKGGYND-NFQGLVQNVRFVGTTPEDILRNKGCSSSTSVLLT 241
: ||| : | : : : ||| : : | : : ||| |||

QY 236 AISENTEITLRGLPHVTTEVGVSSRRPEVCSCEELGNMVOELSGHLVLPNQSENK 295
: ||| : | : : : ||| : : | : : ||| |||
Db 242 L--DNNVNGSSPAIRTNIGHKTLQALICISCELSMVLRLGLRTVITTLQDSIR 299
: ||| : | : : : ||| : : | : : ||| |||

QY 296 RYSDNQELWELIGPPKTRNMSACWQDGRFFAENETWVVDSTCTCTCKFKTKICHQITC 355
: ||| : | : : : ||| : : | : : ||| |||
Db 300 KYTEENKELANELRRPP-----LCYHNGVQYRNEEWTVDSTCTECHCONSVTICKYSC 353
: ||| : | : : : ||| : : | : : ||| |||

QY 356 PPATCASPSFVEGECPCSLHSVDGEGSWPAEWTC 393
: ||| : | : : : ||| : : | : : ||| |||
```

```
Db 354 PIMPCSNATVPDGECCPRC-----WPRC 376

RESULT 12
Q28195 ID Q28195 PRELIMINARY; PRT; 128 AA.
AC Q28195;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE THROMBOSPONDIN-2 (FRAGMENT).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96331130; PubMed=8698834;
RA Lafeuillade B., Pellerin S., Keramidas M., Danik M., Chambaz E.M.,
RA Feige J.J.;
RT "Opposite regulation of thrombospondin-1 and corticotropin-induced
RT secreted protein/thrombospondin-2 expression by adrenocorticotrophic
RT hormone in adrenocortical cells.";
RL J. Cell. Physiol. 167:164-172(1996).
DR EMBL: X89512; CAA61683.1;
DR InterPro: IPR003129; TSPN.
DR Pfam: PF02210; TSPN; 1.
DR SMART: SM00210; TSPN; 1.
DR NON_TER 1
FT NON_TER 128
FT SEQUENCE 128 AA; 14400 MW; 9F0B36F9DA5FA82B CRC64;

Query Match 8.8%; Score 580; DB 6; Length 128;
Best Local Similarity 83.6%; Pred. No. 2.8e-38;
Matches 107; Conservative 12; Mismatches 9; Indels 0; Gaps 0;

QY 23 KDTFTDLSININRKTIGAKQFRGPDGVPAYRFVRFYIPPNADLSKITKIMROKE 82
: ||| : | : : : ||| : : | : : ||| |||
Db 1 EDTAFDLSININRKTIGAKQFRGPDGVPAYRFVRFYIPPNADLSKITKIMROKE 60
: ||| : | : : : ||| : : | : : ||| |||

QY 83 GFFLTALQKDGKSRGTLTLLALEGPGLSQRFIYSNVPADTLDTLYIDGTRHRYVLESDV 142
: ||| : | : : : ||| : : | : : ||| |||
Db 61 GFFLTASMKQDRSRGTLTLLALEGPGATHRQRFIYSNVPADTLDTLYIDGTRHRYVLESDV 120
: ||| : | : : : ||| : : | : : ||| |||

QY 143 GLADSQWK 150
: ||| : | : : : ||| : : | : : ||| |||
Db 121 GLADSQWK 128
: ||| : | : : : ||| : : | : : ||| |||

RESULT 13
Q96RW7 ID Q96RW7 PRELIMINARY; PRT; 5636 AA.
AC Q96RW7;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HEMICENTIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX Trent J.;
RT "Human hemicentin gene.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF156100; AAK68690.1;
DR SEQUENCE 5636 AA; 613660 MW; F000B319CED7B52C CRC64;

Query Match 7.6%; Score 505; DB 4; Length 5636;
```

Db	5164	DCDWTIGSYRCVYRCGSGFRRTSDGLSCODINECOESSPCHQRCFNAIGSFHCCEPGYQ	5223
Qy	902	NDG-----VPDDRDN-CRLVFNPDQEDLDGCRG---DICKDDF---DNDNIPIODDVCP	949
Db	5224	LKGRKCMVNECRQVCR---PDQCKNTRGGYKCIDLCPNGMTKAENGTCIDIDE-CK	5278
Qy	950	EN-----NAISTEDTFNFQMVPLDPKGTQIDPNWVIRHQG-----KEL	988
Db	5279	DGTHQCRNQICENTRSSYRCV-CPRG-----YRSGVGRPCMDIDECEQVPKPC	5327
Qy	989	VQTANSDPGIATGVDFDEFGSVDFTGYVNTDRDDYAGFV---FGYOSSREYVVMWQK	1044
Db	5328	AHQCNSPTG-----SFKICICPPQHLLDGK-SCAGLERLPNIGTYGSS-YNLARFSP	5378
Qy	1045	VTQTYWEDQPTRAYG--YSGVSLKVVNSTT	1072
Db	5379	VRNNYQQQHYROYSHLYSSYS-EYRNSRT	5407
RESULT 14			
ID	Q96SC3	PRELIMINARY; PRT; 2673 AA.	
AC	Q96SC3		
DT	01-DEC-2001 (TrEMBLrel. 19, Created)		
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)		
DE	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)		
DE	FIBULIN-6 (FRAGMENT).		
GN	FIBL-6.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_Taxid=9606;		
RN	SEQUENCE FROM N.A.		
RC	TISSUE=MELANOMA;		
RT	Kostka G., Timpl R.;		
RT	"Partial sequence of fibulin-6 with a c-terminal region related to		
RL	domain II and III of the fibulin family.";		
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AJ306906; CAC37630.1; -.		
FT	NON_TER		
FT	1		
SEQ	SEQUENCE 2673 AA; 291017 MW; BEASC30B8340E272 CRC64;		
Query Match 7.6%; Score 504; DB 4; Length 2673;			
Best Local Similarity 22.3%; Pred. No. 2.6e-30;			
Matches 287; Conservative 147; Mismatches 406; Indels 448; Gaps			
Qy	52	VPAYRFRVFDYIPPVNADLSLTKTKIMRQKEGFFL-TAQLKODGKSRGRTLLALEGPGLSQ	110
Db	1338	IPAH-----FD---SVANGHSELVYRSKSDSYCTTAE-----NSVGFVKAI-----	1378
Qy	111	RQFEIVSNGPADTLDL-TYWID--GTRHVVSLEDVG--LADSW--KNVTVOVAGETYSL	163
Db	1379	-GFYVVKPEPVFGDYPSHWIEPLGNAILNCEVKGDPPTIQMRKGVDEIHSRRQL	1437
Qy	164	HVGCDLDSFALDEPFEYHLQAERSMYVAKGSARESHFRGLLQVHLVFNVEDILSK	223
Db	1438	GNG-----SLAIYGTVNE-AGDYTCVATNEAGVVERSMNLTLSQSPITILEPVEVINA	1491
Qy	224	KG-----CQ-----QGQGA-----EINAISENTETLRGLPHVTTEYVGPSSER	261
Db	1492	GKKIILNCQATGEPQPTITWSRQGHISISWDRVNVLSNN-----SLYI--ADAQ	1538
Qy	262	RPEVCERSCEELGNWQELISGLHVLVYNQPSENLKRVSNDNQFLWELIGPPKTRMMSACW	321
Db	1539	KEDTSEPEC-----VARNLMG-SVLVRVP-----VIVOVHGG-----	1569
Qy	322	QDGRFAENETWVDSCTTCTC-----KKFKTICHQITCP-PATCASP---SFVEGECPCS	373
Db	1570	-----FSQWSAW--RACSV-TCCKGQTKSRSLCNO---PLPANGKPKQGSDELRNRCQN	1618
Qy	374	CLHSVDGEGGSPWAEWTCQSVTCGSGTQQRGRSCDVTS-----NTCLGPSIQTRACSLS	428

```
Db 1619 KPCPVDG--SWSEMSIWEBCRSCGRGNOTRTTCNNFNSVQHGRPCBEGNAVEIIMCNIR 1676
QY 429 KCDTRIDGGHSHWSPSSCSVTCGVGNITRILCNPSVPOMGKNCKGSGRETKACOG 488
Db 1677 PCPVRH---GAWSAWCPMGCTSESCGKGTQTRARLCNPPPAFGSGYCDGAETOMQVCNE 1732
QY 489 APCPIDGRWSPNSACTVTCAGGIRETRVCNSPEQYGGKACVGVQVQERQMCNKRSC 548
Db 1733 RNCPPVHGKATWASASVSCGGGARQTRCSDPVOYGGKRGEGSDVQSDFCNSDPC 1792
QY 549 PVDG-----CLSNPCFPGAQCSSFPD----- 569
Db 1793 PTHGNWSPWSGWTGTCRSCNGSGOMRRTYTCDNPPNSNGRACGPDQSIQRCNTDMCPVD 1852
QY 570 ---GSH-----SCGS-----C-PVGFLNGTHCEDLDECALVPDICFSTSKVP 608
Db 1853 GSWGSHWSPWSOCASCGGKTRKRLCDHPVPVK-GRPCPG-----DTTQVT 1899
QY 609 RCVNTQPGFHCPLPCPPRYRG-----NCPV 632
Db 1900 RC-NVQA---CPGGQRARGSVIGNDVEFGIAFLNATITDSPNSDTRIIPRAKITNVP 1955
QY 633 GVGLEBAK-----TEKQVCEPENCKDKTHNCKHKAECIYLGHFSDPMYKCECQ 681
Db 1956 SLGSAMRKIVSLNPIYWTAKIEGAVNGFT-----LTNAVEKRETO 1998
QY 682 TGYAGDGLI-----CGESD-----LDGWPNLVLCATNATYHIKNCPLHNSG 727
Db 1999 VEFATGEILQMSHIARGLDSDGSLLDIVVSGYV-LQLQSPAETV---VKDY----- 2046
QY 728 QBDPDKDGDGACDDDD---DNDGV-----TDEKDNQC-----LLFNPRQADY 767
Db 2047 TEDIQTGPGQLYATSTRFLTIDGISIPYTNHNTVFDQAQGRMPFLVETLHASSVESDY 2106
QY 768 DKDE-----VGRDNCNCPYVH-----NPAQIDTDNNGEGDAGSDVIDGDDVFN 810
Db 2107 NOIETLGFKIHASISKGDRSNQCPSGFTLDSVGPFCADDECAAGNPCS----- 2156
QY 811 ERDNCPPYVNTQDRTDGGVGDHDCNCP---LVHNPQOT---DVNDLVG-DQCDNNEDI 864
Db 2157 --HSC-----HNMTYYCSPKGLTIAADGRTQDIDECALGRHTCHAGQDC 2202
QY 865 DD-----DGHNQND---NCPYISNANQD---HDR-DQGGD---ACDPDDDD 903
Db 2203 DNTIGSYRCVVRGSGFRRTSGLSCQDINECOESSPCHQRCFNAIGSFHCCEPGYQK 2262
QY 904 G-----VDDDRN-CRLVFNPOEDLDGDRG---DICKDDF---DNDNIPDIDVCPEN 951
Db 2263 GRKMDVNECRQNVCR---PQHCKNTRGGYKCIDLCPNGMTKAENGTCIDIDE-CKDG 2317
QY 952 -----NAISETDRNFQVPLDPKGTQIDPNWVIRHQG-----KELVQ 990
Db 2318 THQCRYNQICENTRSYRCV---CPRG-----YRSGVGRPCMDINEQVFKPAH 2366
QY 991 TANSDPGIAVGDFEGSVDFSTFTVNTDRDDYAGFV-----FGYOSSRFYVVMKQVT 1046
Db 2367 QCSNTPG-----SFKICPPGQHLLDGK---SCAGLERLPNGYQYSS-YNLARFSPVR 2417
QY 1047 QYWEQDPTRAYG--YSGVSLAVNSTT 1072
Db 2418 NNYQPOOHYRQYSHLYSSYS-EYRNSRT 2444
```

```
RESULT 15
Q9TTS5 PRELIMINARY: PRT: 685 AA.
AC Q9TTS5;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE SCO-SPONDIN (FRAGMENT).
GN SCO-SPONDIN.
```

```
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-SUBCOMMISSURAL ORGAN;
RA Gobron S., Creveaux I., Monnerie H., Elbitar F.;
RT "Characterization of cattle SCO-spondin.";
RL Submitted (Oct-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ132106; CAB53759.1;
DR InterPro; IPR002919; TIL.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF01826; TIL; 2.
DR Pfam; PF00090; TSP1; 6.
DR SMART; SM00209; TSP1; 6.
DR PROSITE; PS00092; TSP1; 4.
FT NON_TER 1
FT NON_TER 685
SQ SEQUENCE 685 AA; 71389 MW; A003F880E746D5D3 CRC64;

Query Match 6.7%; Score 442; DB 6; Length 685;
Best Local Similarity 24.5%; Pred. No. 2.8e-26;
Matches 125; Conservative 39; Mismatches 146; Indels 200; Gaps 19;

Qy 320 CWQDGRFAENETWVVDCTTCKKFKTICHQIT-CPPATCASPFEVEGECPCSLHVS 378
Db 227 CTGGRLLHPPGSA-VLRPCENCSC-----VSLITNCTSWPK----- 263
Qy 379 DGEWGSPAWEWTCQSVTCGSGTQQRGSC-----DVTSTNCTLGPSTQ 421
Db 264 EQQPTSPWTPWSECSASCGPARRKHFRCTRPPGAPSSMAPLLSSVPPLCPEAE 323
Qy 422 TRACLSKCDTRIRQDGGSHWSPSSSVTCGVGNITRILCNPSVPOMGKNCKGSR 481
Db 324 EEPCLLPECD---RAGMGPMGWSPSSCSRSCGGLRSRACDQPPPGGLDGYCEGPRA 379
Qy 482 ETAKCQ----- 487
Db 380 QGAACQALPCPVNTCTATEGAEYSACGPPCRSCDDLVHCVHWCQPCYCPGQVLSADG 439
Qy 488 -----GA-----PCPIDGRWSPWSP 502
Db 440 TVHVQPGHCSCLDLTLGERHRPGAQLAKPDGCNYCTCSEGLTCTDLPCVPFGAWCPWSE 499
Qy 503 WSACTVTTCAGGIRETRVCNSPEQYGGKACVGVQVQERQMCNKRSCPVDDGCLS 555
Db 500 WTACSPQCGQGTTRSRACSCAPQOHGAPCPGEGAGEAGAHQRETCASTPECPVDGAW 559
Qy 556 NPCFPGAQCS-----SFPDGSWSCGSCPVGF-----LGNGTHCEDLDECALV 597
Db 560 -PWGSPWSEVCLGRSHRSRECSWPTSEGGRCPPGHRQSRPCOGNSTQCTD---CAGG 615
Qy 598 PDICFSTSKVPR-CVNTQPGFHCPLCPPRYRGNQVGVGLEAAKTEKQVCEPENCKDKT 656
Db 616 QDILLPCGQPCPRSCDLSPGVEEC-----QPSMSGCGQQPRCG---CPEGQLSDGL 662
Qy 657 HNCXKHAECIYLGHFSDPMYKCECOTGYAG 686
Db 663 --CVTPSQC-----RCQYQPGAMG 679

Search completed: August 9, 2002, 10:06:10
Job time: 291 sec
```